Astrophysical Tau Neutrinos in IceCube

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)

im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Juliana Stachurska, M.Sc.

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Marek Kowalski, Humboldt-Universität zu Berlin.
2. PD Dr. rer. nat. Walter Winter, Humboldt-Universität zu Berlin.
3. Prof. Dr. Claudio Kopper, Michigan State University, MI, USA.

Tag der mündlichen Prüfung: 30. April 2020

Abstract

The IceCube neutrino observatory at the South Pole has confirmed the existence of a diffuse astrophysical neutrino flux and measured it in multiple channels. While the first extragalactic neutrino source has likely been identified, the sources of the majority of astrophysical neutrinos remain unknown. The flavor composition of astrophysical neutrinos carries information on the environments at the sites of cosmic particle acceleration as well as potential imprints of new physics acting during neutrino propagation. To tightly constrain the flavor composition the observation of the long-elusive tau neutrinos is required. Starting at an energy of $\sim \mathcal{O}(100\text{TeV})$ a tau neutrino charged current interaction can produce a double cascade topology, where the two energy depositions from the tau creation and the tau decay vertices are resolvable in IceCube. This topology together with the well-established track and single cascade topology is used to measure the flavor composition on Earth. In this work, high-energy events starting in IceCube's detector volume are classified algorithmically into the three topologies. In the dataset with a livetime of 7.5 years, two events are classified as double cascades for the first time, yielding multi-TeV tau-neutrino candidates. Their observation is consistent with the expectation, assuming astrophysical neutrinos and standard neutrino oscillations. The reconstructed properties of the two tau-neutrino candidates are investigated in more detail in an a-posteriori analysis, making use of targeted Monte-Carlo simulation. The statistical method to constrain the flavor composition is improved by performing a log-likelihood-ratio test using multi-dimensional probability densities. The probability of each of the tau-neutrino candidates to stem from a tau-neutrino interaction versus any other scenario is assessed. One of the double cascades is consistent with being a misclassified single cascade, while the second double cascade is found to have a misclassification probability of only 3%. A new flavor composition measurement is performed using the new multi-dimensional likelihood. The measured flavor composition $\nu_e : \nu_\mu : \nu_\tau = 0.20 : 0.39 : 0.42$ is consistent with astrophysical neutrinos from all possible astrophysical production mechanisms, as well as with all previously published results. The astrophysical tau-neutrino flux is measured to $\frac{\mathrm{d}\Phi_{\nu_\tau}}{\mathrm{d}E_\nu} = 3.0^{+2.2}_{-1.8} \cdot \left(\frac{E}{100\ \text{TeV}}\right)^{-2.87} \cdot 10^{-18} \cdot \text{GeV}^{-1}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}$ with spectral index $\gamma = 2.87^{+0.21}_{-0.20}$, yielding the first non-zero results for the tau normalization. The absence of an astrophysical tau-neutrino flux is disfavored at 2.8σ.

Zusammenfassung

Das IceCube Neutrino Observatorium am Südpol hat die Existenz eines diffusen astrophysikalischen Neutrinoflusses nachgewiesen und ihn in mehreren Kanälen mit hoher Signifikanz gemessen. Obwohl die erste wahrscheinliche extragalaktische Neutrinoquelle identifiziert wurde, bleiben die Quellen der Mehrheit der astrophysikalischen Neutrinos unbekannt. Die Flavor-Zusammensetzung astrophysikalischer Neutrinos trägt Informationen sowohl über die Orte kosmischer Teilchenbeschleunigung, als auch über die Auswirkungen potenzieller neuer Physik auf die Neutrinoausbreitung. Zur Bestimmung der Flavor-Zusammensetzung ist die Beobachtung der lange nicht fassbaren Tau-Neutrinos von Nöten. Ab einer Energie von $\sim \mathcal{O}(100\text{TeV})$ kann die Wechselwirkung eines Tau-Neutrinos über geladene Ströme eine Doppelkaskaden-Topologie ergeben, bei der die zwei Energiedepositionen am Tau-Entstehungsvertex und dem Tau-Zerfallsvertex in IceCube aufgelöst werden können. Diese Topologie wird zusammen mit den bereits bekannten Topologien einer Einzel-Kaskade und einer Spur zur Messung der Flavor-Zusammensetzung auf der Erde benutzt. In dieser Arbeit werden im Detektorvolumen von IceCube anfangende Ereignisse mit hohen Energien algorithmisch in die drei Topologien klassifiziert. Im Datensatz mit einer Lebensdauer von 7.5 Jahren werden zum ersten Mal zwei Ereignisse als Doppelkaskaden klassifiziert; diese sind Kandidaten für multi-TeV Tau-Neutrinos. Deren Beobachtung entspricht den Erwartungen von astrophysikalischen Neutrinos und standard Flavor-Oszillationen. Die rekonstruierten Eigenschaften der zwei Tau-Neutrino-Kandidaten werden mithilfe gezielter Monte-Carlo Simulation in einer a-posteriori Analyse im Detail studiert. Die statistische Methode der Bestimmung der Flavor-Zusammensetzung wird durch einen Log-Likelihood-Quotienten-Test mit multi-dimensionalen Wahrscheinlichkeitsdichten verbessert. Die Wahrscheinlichkeit der zugrunde liegenden Wechselwirkung eines Tau-Neutrinos wird für jeden der Tau-Neutrino-Kandidaten gegen andere Szenarien verglichen. Eine der Doppelkaskaden ist konsistent mit dem Szenario einer misklassifizierten Einzelkaskade, während für die zweite Doppelkaskade die Wahrscheinlichkeit eines nicht-Tau-Neutrino Szenarios auf nur 3% bestimmt wird. Mit der neuen multi-dimensionalen Likelihood wird eine neue Messung der Flavor-Zusammensetzung durchgeführt. Die gemessene Flavor-Zusammensetzung ist konsistent mit der Annahme von astrophysikalischen Neutrinos aller astrophysikalischer Produktionsmechanismen, wie auch mit bisher veröffentlichen Resultaten. Die Messung ergibt einen astrophysikalischen Tau-Neutrino Fluss von $\frac{\mathrm{d}\Phi_{\nu_\tau}}{\mathrm{d}E_\nu} = 3.0^{+2.2}_{-1.8} \cdot \left(\frac{E}{100\ \mathrm{TeV}}\right)^{-2.87} \cdot 10^{-18} \cdot \mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ mit spektralem Index $\gamma = 2.87^{+0.21}_{-0.20}$, was dem ersten positiven Ergebnis für die Tau-Normalisierung entspricht. Die Nichtexistenz eines astrophysikalischen Tau-Neutrino Flusses wird mit einer Signifikanz von 2.8σ abgelehnt.

Two tau or not two tau,
that is the question.

Contents

Chapter 1

Introduction

The era of astroparticle physics began in 1912, when Victor Hess observed an increase in ionizing radiation as a function of altitude during balloon flights [1], contradicting the then-held belief that the radiation was produced inside the Earth. The only possible explanation was the correct one, namely that the radiation had a cosmic origin. Thus, the ionizing radiation was named cosmic rays. During the next century, these cosmic rays were revealed to be high-energy charged particles, mainly protons, that penetrate the Earth's magnetic field. Their flux spans over 10 orders of magnitude in energy and can be described by power laws. It has been studied in great detail, revealing spectral shape features such as the "knee" and the "ankle". At the highest energies, the energetics and largely isotropic arrival directions require an extragalactic origin of cosmic rays. Being charged, cosmic rays are deflected by magnetic fields en-route. They consequently do not point back to their sources, which thus remain unknown. Astroparticle physics aims to understand these sources and the mechanisms leading to the production of the high-energy cosmic rays observed on Earth.

The universe has been extensively studied in the electromagnetic spectrum, as its messengers – photons – travel in straight lines. Many source classes are known to produce very high-energy gamma rays. Photons are always produced at sites of particle acceleration, both in collisions of cosmic rays and via the radiative energy loss processes of high-energy charged particles. It is thus natural to suspect that the gamma-ray sources might also produce the diffuse cosmic-ray flux measured on Earth. However, photon production does not necessarily imply the simultaneous acceleration of charged hadrons to very high energies. Further, the universe is opaque to the highest-energy photons. Yet another particle can potentially help solve the mystery of the origin of cosmic rays. This particle is the neutrino, postulated in 1930 by Wolfgang Pauli [2]. Shortly after,

Enrico Fermi explained nuclear beta decay as a neutron decaying into a proton, an electron, and an anti-electron neutrino [3]. He also named the neutrino, "the little neutral one", and made the first attempt to measure its mass using the endpoint of the beta-decay electron spectrum. He arrived at the conclusion that the rest mass of neutrinos must be either zero or orders of magnitude below the electron's rest mass. To date, only upper limits on the absolute neutrino mass scale exist [4, 5]. Beta decay electrons are still employed in present [5] and future [6] neutrino mass experiments.
The electron neutrino remained elusive until it was first observed two decades later by Clyde Cowan and Frederick Reines [7]. Like the charged leptons, neutrinos come in three flavors. If a neutrino interacts via charged current, it transforms into its corresponding charged lepton. This forms the basis of identifying the interacting neutrino's flavor. The muon neutrino was observed in 1964 by Lederman, Schwartz and Steinberger [8]. Due to the high mass and short lifetime of the tau, the tau neutrino took much longer to detect, and was finally observed in 2000 by the DONUT collaboration [9]. The tau neutrino has remained the most elusive particle of the standard model: only two experiments were able to positively identify tau-neutrino interactions. DONUT observed nine and OPERA four [10], yielding a grand total of 13 previously identified tau-neutrino events.
In the meantime, the standard model of particle physics was developed, in which neutrinos were considered massless particles that only interact weakly. However, the observation of solar neutrinos provided the first clues that something was missing: the calculated solar neutrino flux [11] was higher than the observed one [12]. Decades later, the problem was solved by neutrino oscillations [13, 14]. However, the accepted model of neutrino oscillations requires tiny differences between the neutrino masses, immediately implying that the neutrino mass is actually non-zero. This was the first experimental confirmation calling for an extension of the standard model.
Neutrinos can be seen as ideal cosmic messengers. Stable, almost massless, with no electric charge and thus interacting only weakly, they can reach us from their sources in the distant universe without deflection or absorption. By precisely reconstructing their direction, and accumulating a large number of astrophysical neutrinos, neutrino sources should be resolved. Neutrinos themselves cannot be accelerated at the sources, their production requires the acceleration of charged particles, which can escape the sources and be observed on Earth as cosmic rays. Neutrino sources are thus cosmic-ray sources. Their small interaction probabilities make them difficult to detect, requiring enormous detector volumes in low-background environments. The history of neutrino astronomy [15] goes back to 1961, when Moisey Markov put forward the idea for a cubic-km-scale subsurface detector [16]. Based on this idea, the *Deep Underwater Muon And Neutrino Detector*

(DUMAND) project was founded. It envisioned the deployment of optical sensors deep in the ocean off the shore of Hawaii to detect muons from atmospheric and astrophysical neutrinos. The deep location would provide shielding against the much more abundant muons produced in cosmic-ray interactions in the atmosphere [17]. Although the project was cancelled after the installation of only one string, the advances in technology were applied to the smaller-scale *Baikal Neutrino Telescope* that was successfully installed in Lake Baikal in Russia [18] and observed high-energy atmospheric neutrinos [19]. The experiment *Astronomy With A Neutrino Telescope And Abyss Environmental Research* (ANTARES) is currently operating a similar detector in the Mediterranean Sea, while second-generation detectors are currently under construction at both locations.
Francis Halzen and John Learned devised a detector in the ice [20], which would develop into the *Antarctic Muon And Neutrino Detector Array* (AMANDA) [21], operating at the South Pole until 2009, and serving as a precursor to IceCube.
The *IceCube Neutrino Observatory* [22] is the first realization of a cubic-km neutrino detector and was completed in 2010. It consists of more than 5000 optical sensors deployed into the Antarctic ice close to the geographic South Pole. In 2013, IceCube reported the first observation of astrophysical neutrinos [23]. Later measurements confirmed the presence of astrophysical neutrinos in different channels [24, 25, 26] and characterized the flux. The measured astrophysical flux is consistent with following a single powerlaw, and with being made up of electron, muon and tau neutrinos in equal amounts. No significant self-clustering of neutrinos has been observed, and the lack of galactic clustering suggests a predominantly extragalactic origin. However, one high-energy neutrino has been associated with a likely source, an active galaxy which at that time happened to be in a state of highly enhanced electromagnetic activity. The identification of TXS 0506+056 as the first likely high-energy neutrino and cosmic-ray source was a breakthrough in the young field of multi-messenger astronomy. TXS 0506+056 marks only the third known extraterrestrial neutrino source, after our Sun [12] and supernova 1987A [27, 28, 29].
The flavor composition of astrophysical neutrinos carries information about the processes leading to high-energy neutrino production at the sources. It can thus help to identify leading source classes of neutrinos and therefore of cosmic rays, even if the source class dominating neutrino production is distributed too diffusely to be resolved. Various conceivable neutrino production scenarios lead to differing source flavor compositions, none of which contains a significant fraction of tau neutrinos: $\nu_e : \nu_\mu : \nu_\tau = x : 1-x :\sim 0$ with $0 \leq x \leq 1$ at the sources. Due to the long distances that the astrophysical neutrinos travel, neutrino oscillations will average out, and become distance-independent. For any

realization of the source flavor composition, we expect astrophysical neutrinos to arrive at Earth with a flavor composition with fractions $0.15 < f_{\nu_e} < 0.6$, $0.2 < f_{\nu_\mu} < 0.45$, and $0.15 < f_{\nu_\tau} < 0.5$, assuming three-flavor oscillations [30]. An astrophysical tau-neutrino flux is thus guaranteed, and its absence would implicate new physics. At the energies of $\mathcal{O}(10)$ Tera electron Volt (TeV)$-\mathcal{O}(10)$ Peta electron Volt (PeV) where the astrophysical neutrino flux is measured, the atmospheric backgrounds for tau-neutrinos are negligible. Each identified high-energy tau neutrino is thus an astrophysical neutrino. A measurement of a pure sample of tau neutrinos would not only provide a background-free dataset for neutrino source searches, but would also allow for an independent measurement of the parameters of the astrophysical neutrino spectrum. Owing to the great distances between the neutrino sources and Earth, the measured neutrino flavor composition is also sensitive to new physics scenarios affecting neutrino propagation or interaction [31]. The difficulty in identifying tau-neutrino charged-current interactions, combined with the low number of observed high-energy neutrino events in IceCube, explains why no astrophysical tau-neutrino has thus far been identified despite dedicated searches.

To identify the interacting neutrino's flavor, it is of utmost importance to understand the topologies that can be created by neutrinos interacting in IceCube. Muons passing through the ice leave long tracks of Cherenkov light. These track-like events are predominantly created by charged-current muon-neutrino interactions, but also stem from atmospheric muons and from the 17% of charged-current tau-neutrino interactions in which the tau decays to a muon. Cascade-like events are created by approximately spherically symmetric Cherenkov emission from a single stationary light source. Charged-current electron-neutrino interactions and the majority of charged-current tau-neutrino interactions create cascades. In addition, all neutrino flavors create cascade-like events when interacting via neutral current. These two topologies were used for the first flavor composition measurements of astrophysical neutrinos, yielding best fits of $\nu_e : \nu_\mu : \nu_\tau = 0.5 : 0.5 : 0.0$ [32] and $\nu_e : \nu_\mu : \nu_\tau = 0.0 : 0.2 : 0.8$ [33]. These two results are in agreement, owing to their large uncertainties, which in turn are given by the limited sensitivity to constrain three flavors with a binary topology classification. A third topology is necessary to break the degeneracy.

John Learned and Sandip Pakvasa realized that a tau with energy above 1 PeV may live long enough to travel from one optical sensor of a segmented neutrino detector to another. It can then create a so-called "double bang", consisting of two energy depositions connected by a particle traveling with the speed of light. The tau neutrino interaction produces the first cascade-like energy deposition (or "bang"), and the tau decay produces the second "bang" if the tau does not decay into a muon. 83% of charged-current

tau-neutrino interactions follow this double-bang channel, but the vast majority are not distinguishable from electron-neutrino charged-current interactions due to the tau's short lifetime. In IceCube, the spacing of the optical sensors requires a multi-PeV tau-neutrino to create a double bang that is resolvable by eye. Following the first dedicated search for high-energy tau neutrinos in IceCube [34], improved tau-neutrino detection methods have been developed, lowering the energy threshold. If the tau-neutrino interaction and subsequent tau decay happen close to a sensor, a signature called a "double pulse" can be observed [35] in that sensor, stemming from light from each of the vertices arriving at different times. The approach taken here uses a dedicated algorithm to reconstruct the topology of two causally-connected cascades [36]. This topology is called a "double cascade" and can be resolved down to several meters of tau propagation lengths. By differentiating between single cascades and double cascades, the degeneracy between ν_e and ν_τ interactions is broken for high-energy neutrinos starting at energies $\sim \mathcal{O}(100)$ TeV. Classifying events into single cascades, double cascades and tracks, the contributions of each of the three flavors to the astrophysical neutrino flux can be inferred. The first analysis using the three topologies and an algorithmic topology classification based on the events' observables was developed for IceCube's high-energy starting events with six years of livetime. No double cascades were identified, leading to a measured flavor composition on Earth of $\nu_e : \nu_\mu : \nu_\tau = 0.5 : 0.5 : 0.0$ and an upper limit on the astrophysical tau-neutrino flux of $\Phi_{\nu_\tau}(E_{\nu_\tau}) < 2.95 \cdot 10^{-18} (E_{\nu_\tau}/100\ \text{TeV})^{-2.94}\,\text{GeV}^{-1}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}$ [37].

In the work presented in this thesis, an updated tau-neutrino search and flavor composition measurement using high-energy starting events in IceCube is presented. An additional 1.5 years of livetime are added, but also the previous analysis is improved upon. The algorithmic ternary topology classification is incorporated into the sample of high-energy starting events, an improved likelihood is used for the fitting, the treatment of systematic uncertainties is updated, and previously taken data are reprocessed following an improved detector calibration. As a result, two events are reclassified from single cascades to double cascades. The statistical treatment of the double cascade observables is improved upon, based on targeted Monte Carlo simulations for individual events. An unbiased method to evaluate sparse data in multiple dimensions is employed to properly evaluate all properties of the double cascades that are sensitive to neutrino flavor. The flavor composition is measured using a combined unbinned likelihood for the double cascades and a binned likelihood for the single cascades and tracks, in a maximum-likelihood multi-component fit.

This thesis is organized as follows. In Chapter 2, neutrino astroparticle physics is reviewed. Further, neutrino production, propagation and interaction are introduced and atmospheric and astrophysical neutrinos are described. The IceCube Neutrino Observatory is described in Chapter 3, concentrating on the detector components and neutrino detection principle, as well as the Monte Carlo generation of neutrino and muon events and event reconstruction methods. Chapter 4 is dedicated to high-energy neutrino topologies and the high-energy starting event selection. In Chapter 5, the ternary topology classification and the flavor composition measurement are explained. In Chapter 6, the initial results are presented and the observables of the two found double cascades are discussed. In Chapter 7, the a posteriori analysis of the tau-neutrino candidates is presented. The targeted Monte Carlo simulation is described, and the multi-dimensional statistical treatment is developed. The final results are presented and discussed. The thesis is summarized and an outlook is given in Chapter 8.

Chapter 2

Neutrino Astroparticle Physics

Neutrinos are fundamental particles. They are stable, have no charge and almost no mass and only interact weakly. They do not get deflected or absorbed during propagation and thus can reach Earth from the distant universe. In this chapter, the basics of neutrino astroparticle physics will be reviewed. Cosmic rays are reviewed in Section 2.1, followed by a description of particle acceleration mechanisms at astrophysical sources that can lead to an observable neutrino signal on Earth in Section 2.2. What happens to neutrino flavor during neutrino propagation is explained in Section 2.3, followed by neutrino interactions at high energies in Section 2.4. Finally, in Section 2.5 the astrophysical neutrino flux and in Section 2.6 the atmospheric neutrino and muon fluxes observed by IceCube are introduced.

2.1 Cosmic Rays

The Earth's atmosphere is constantly hit by cosmic ray (CR) particles. These are fully ionized nuclei, about 90% protons, 9% alpha particles, and a small fraction of heavier nuclei up to iron [39]. Figure 2.1 shows the cosmic ray spectrum as measured by air shower experiments. The cosmic ray flux can be described by a power law as $\mathrm{d}N/\mathrm{d}E \sim E^{-\gamma}$ over ~ 10 orders of magnitude, with spectral breaks called the "knee", "second knee", and "ankle" [38]:

- $\gamma \approx 2.7$ for 10^{10} eV $\lesssim E \lesssim E_{\mathrm{knee}} \sim 10^{15-16}$ eV,
- $\gamma \approx 3.1$ for $E_{\mathrm{knee}} \lesssim E \lesssim E_{\mathrm{ankle}} \sim 10^{18.5}$ eV,
- $\gamma \approx 2.6$ for $E_{\mathrm{ankle}} \lesssim E \lesssim E_{\mathrm{cutoff}} \sim 10^{19.5}$ eV.

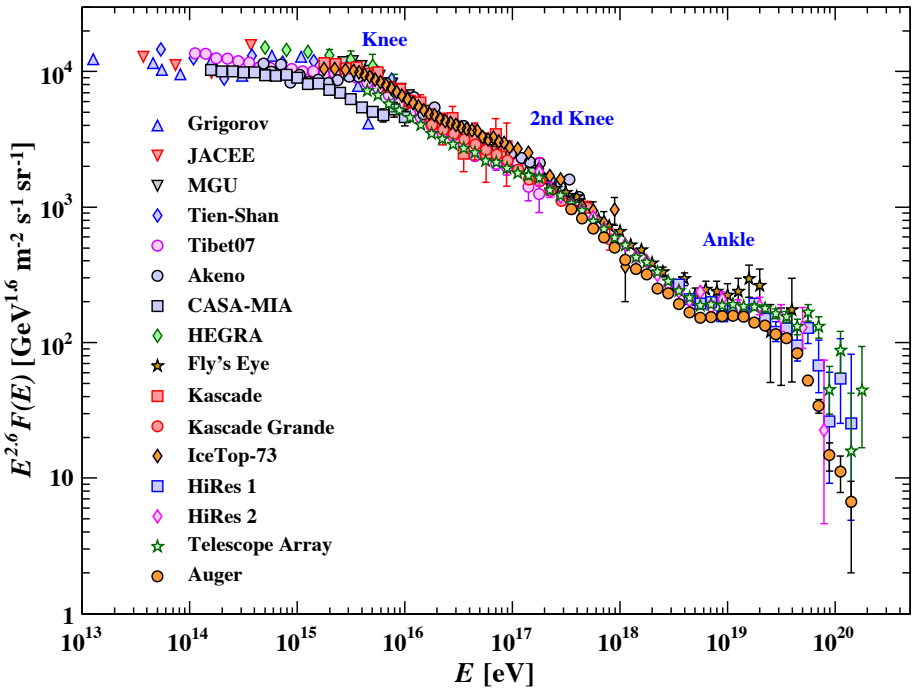

Figure 2.1: The all-particle cosmic-ray spectrum as a function of energy per nucleus measured by air shower experiments. Figure from [38].

The "second knee" at $E_{\text{2nd knee}} \sim 8 \cdot 10^{16}$ GeV is a subtle feature, where the spectrum slightly softens. Above $E \sim 10^{19.5}$ GeV, the spectrum shows a cutoff, which could be due to the Greisen-Zatsepin-Kuzmin (GZK) cutoff, i.e. caused by interactions of ultrahigh energy cosmic rays (UHECR) with the photons of the cosmic microwave background (CMB) [40, 41]. Data from the Pierre Auger Observatory indicate the composition getting heavier at energies $E > 10^{18}$ GeV [42, 43], i.e. the fraction of heavy nuclei increases. In this case, the cutoff is compatible with the effects of photodissociation of heavy nuclei. At the lowest energies, the CR cannot penetrate the Earth's shielding magnetic field. At about 10 GeV, about 10^4 m^{-2} s^{-1} cosmic rays hit the atmosphere, while at the highest energies, the flux is ~ 0.001 km^{-2}yr^{-1}

Most of the cosmic rays are of galactic origin, some are produced by our own Sun, but at the highest energies, cosmic rays come from sources yet unknown and from beyond our galaxy. The acceleration of the cosmic rays at their sources should also lead to neutrino production, as will be explained in the next section. As CR are ionized, massive and charged particles, they get deflected during propagation through galactic and intergalactic magnetic fields. Therefore, neutrinos can greatly add to our understanding of cosmic rays, their origins and their production mechanisms.

Up to energies of $\sim 10^{14}$ eV, cosmic rays can be measured with satellite based instruments. Using calorimetric detectors, the primary particle can be identified. However, above 100 TeV, the low cosmic-ray flux requires larger detector areas than can be comfortably fit on satellite missions.
Upon impacting the atmosphere, CR collide with nuclei in the atmosphere, producing cascades of secondary particles. The processes are not unlike processes happening in beam dumps at particle accelerators on Earth. The secondary particles, mainly pions, kaons, but also heavier charmed hadrons, interact and decay, finally leading to the particles that are observed with ground-based detectors. At very high energies, cosmic rays create extensive air showers, which can be studied with large and sparse detector arrays such as the Pierre Auger Observatory [44] or the Telescope Array [45], making use of their large footprints on the ground. When high-energy charged particles from the shower reach the ground, they can be detected via the Cherenkov radiation they induce in water-filled tanks as employed by the Pierre Auger Observatory, or via scintillation light induced in scintillator surface arrays as employed by the Telescope Array. Both observatories also have detectors that capture the fluorescent light from nitrogen atoms in the atmosphere, that have been excited by charged secondary particles in the shower. The shower development can be calculated using coupled cascade equations, describing the losses and decays of the particles in the shower [39, 46]. The cascade equations use hadronic models, which describe the interactions of the particles in the cascade. As the hadronic models are tuned to accelerator data, they need to be extrapolated to the UHECR regime, introducing a source of systematic uncertainties. Among all the particles produced, only muons and neutrinos can reach subsurface detectors such as IceCube. It is worth mentioning that the muon multiplicity as observed by the Pierre Auger Observatory is in tension with predictions from all hadronic models [47, 48].

2.2 Cosmic Neutrino Sources and Production Mechanisms

As will be described in this section, neutrino production mechanisms also lead to the production of CR and photons. Thus, when trying to explain the origin of high-energy astrophysical neutrinos, a natural start are sources known to produce high-energy photons. Many source classes have been proposed to explain the diffuse astrophysical neutrino flux observed by IceCube. However, the analyses performed with IceCube thus far only yielded upper limits on the contribution of the tested source classes to the observed neutrino flux. The observation of sources within our own galaxy in very-high-energy gamma rays makes a galactic contribution to the astrophysical neutrino flux likely. E.g.

supernova remnants (SNR) are known to be good particle accelerators. However, recent searches for correlations of IceCube neutrinos with the galactic plane or known SNR found no significant correlations and limit the galactic contribution to 14% [49], while another recent analysis revealed hints of a diffuse galactic contribution at the 2σ level [50]. Most of the diffuse astrophysical neutrino flux thus seems extragalactic in origin. A former favorite source class, Gamma-Ray-Bursts (GRBs) have been shown to contribute less than 1% to the diffuse neutrino flux [51]. Blazars, jetted Active Galactic Nuclei (AGN) with the jet pointed at Earth, were limited to a maximum of 27% of the diffuse flux [52], and supernovae to a maximum of 13% in the choked-jet scenario and 26% in the Type IIn scenario [53]. More recently, Tidal Disruption Events (TDEs) have been proposed as neutrino source candidates, also their contribution has been limited to 26% for non-jetted and 1.3% for jetted TDEs [54]. While source classes have been proposed and subsequently disproven, and also searches for self-clustering of IceCube neutrinos have come up empty-handed[1], a search for electromagnetic counterparts to high-energy muon neutrinos has been more fruitful. On September 22, 2017, a high-energy neutrino triggered the realtime alert system [56]. The IceCube-170922A alert [57] was followed up by 16 observatories. In the 50% containment region of the neutrino, *Fermi* reported the blazar TXS 0506+056 to be in a flaring state [58], and MAGIC reported very high-energy gamma rays from this blazar [59]. This was the first compelling evidence for an electromagnetic counterpart to a high-energy IceCube neutrino [60]. In an independent archival search for neutrino emission from the same position, an excess of neutrinos was found which was not coincident with enhanced electromagnetic emission [61]. While the archival neutrino flare further strengthened the case of TXS 0506+056 as a neutrino source, it also poses a problem for the modeling of particle acceleration in blazars [62].

2.2.1 Particle Acceleration at Sources

As neutrinos are electrically neutral, they cannot be accelerated in sources. Instead, they are created in hadronic processes when charged particles interact with the matter or radiation fields at the sources that accelerate them. These interactions create secondary particles, whose interactions and decays create neutrinos. The primary accelerated particles eventually escape the source and may reach Earth as cosmic rays. Thus, neutrino sources are also cosmic ray sources. Finding cosmic neutrino sources able to produce neutrinos with energies up to an energy range of Exa electron Volts (EeV) would at the

[1] The latest point-source search for steady neutrino emission from known astrophysical sources revealed several sources that might become statistically significant in the next few years, among them TXS 0506+056. Combined, the significance of the hottest four sources, each of them with $< 3\sigma$ significance, is 3.3σ after correcting for trial factors [55].

same time yield some answers to the 100-year-old questions of the origin, the sources and the acceleration mechanisms of the highest-energy cosmic rays. While cosmic rays get deflected by magnetic fields during propagation, neutrinos and photons do not and thus point back to their sources. However, photons are more easily absorbed, and can further stem from leptonic processes, thus high-energy gamma-ray sources do not automatically need to also be cosmic ray sources.

One method to accelerate particles to very high energies at cosmic accelerators is **first order Fermi acceleration**, also called **diffusive shock acceleration**. The description below is based on [39]. If a shock propagates through a medium with a velocity $u_1 \ll c$, then in the frame of the shock, the gas upstream of the shock enters the shock with velocity u_1, while downstream of the shock the medium is dragged behind and departs the shock with velocity u_2. A relativistic particle can enter the shock, and scatter elastically in the shocked medium, until it leaves the shock again. By moving across the shock front twice, it can gain energy. The average gain in energy per crossing cycle is proportional to the velocity difference across the shock, and averaging over directions, $\langle \Delta E/E \rangle = \langle \delta(E) \rangle \approx 4/3(u_1 - u_2)/c$. After n cycles of crossing the shock front back and forth, the particle has the energy $E_n = E_0(1+\delta)^n$. But the particle also has an escape probability P_{esc} from the accelerating region, thus the probability to be accelerated over n cycles is $(1 - P_{\text{esc}})^n$. P_{esc} depends on the difference in velocities between the shock front and the medium downstream, $P_{\text{esc}} = 4u_2$. Thus the number of particles that can be accelerated to at least an energy E_n is, using $n = \log(E_n/E_0)/\log(1+\delta)$,

$$\begin{aligned} N(E \geq E_n) &\approx N_0(1 - P_{\text{esc}})^n \\ &= N_0(E_n/E_0)^{\log(1-P_{\text{esc}})/\log(1+\delta)} \\ &= N_0(E_n/E_0)^{1-\gamma}, \end{aligned} \tag{2.1}$$

where $\gamma = 1 - \log(1 - P_{\text{esc}})/\log(1+\delta) \approx 1 + P_{\text{esc}}/\delta = 1 + 3/(u_1/u_2 - 1)$ is introduced. u_1 and u_2 are related by mass conservation at the shock and, assuming an ideal gas ($c_p/c_v = 5/3$) [63], one obtains:

$$\begin{aligned} u_1/u_2 &= \frac{c_p/c_v + 1}{(c_p/c_v - 1) + 2/M^2} \\ &\approx 4(1 - 3/M^2) \end{aligned} \tag{2.2}$$

with the Mach number $M = u_1/c_s$, where c_s is the speed of sound. Using the above expression, one finds $\gamma \approx 2 + 4/M^2$. This naturally gives the differential power spectrum

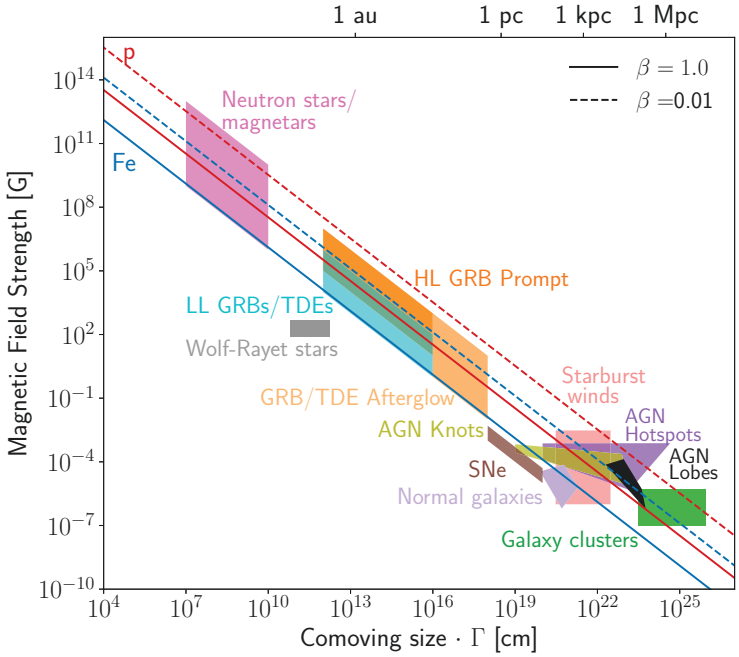

Figure 2.2: Hillas diagram. Source classes are shown according to their size and magnetic field strength. The diagonal lines indicate the threshold for BR beyond which the source can confine protons (red) and iron nuclei (blue) with energies of 10^{20} eV for shock velocities $\beta = u_1/c$. Sources above the diagonal line satisfy the Hillas criterion. Figure from [64].

$\mathrm{d}N/\mathrm{d}E \approx NE^{-2}$ for strong shocks with high Mach number, where N is now a normalization factor fixed to some energy.

A source can accelerate particles while their gyroradius is much smaller than the size of the acceleration region. The maximum energy a particle can reach before it escapes the accelerating region is given by the Hillas criterion [65]:

$$\frac{E_{\mathrm{max}}}{10^{18}\mathrm{eV}} = \frac{3}{20}\frac{u_1}{c} Z \left(\frac{B}{\mathrm{G}}\right)\left(\frac{R}{10^{15}\mathrm{m}}\right) \tag{2.3}$$

where Z is the charge of the particle, B is the source's magnetic field, and R is the radius of the accelerating region. The so-called Hillas diagram in Figure 2.2 shows source classes in terms of their radial size and magnetic field strength. Based on the radius and magnetic field considerations alone, normal galaxies such as the Milky Way cannot produce UHECR. Very large objects such as galaxy clusters or objects with extremely high magnetic fields such as neutron stars, however, can in principle confine 10^{20} eV protons. Fulfillment of the Hillas criterion alone does not guarantee cosmic-ray acceleration to

the highest energies.

There are a few main scenarios of how to produce neutrinos from the accelerated cosmic rays at sources. In the hadronuclear scenario, cosmic rays, which are predominantly protons, interact with surrounding matter, which mostly consists of hydrogen. This is also called the pp-scenario (and if neutrons are involved, the pn-scenario). In this scenario, the interaction of high-energy cosmic rays with the target nuclei creates a particle shower, similar to the particle shower created by cosmic-ray interactions in the Earth's atmosphere. The most abundant particles created in those showers whose decay leads to neutrino production, are charged pions, but there are also contributions from kaons and heavier mesons. In the photohadronic, or pγ-scenario, the cosmic rays interact with radiation, and pions are predominantly created via the Δ-resonance:

$$p + \gamma \to \Delta^+ \to \begin{cases} p + \pi^0 & (2/3 \text{ of cases}), \\ n + \pi^+ & (1/3 \text{ of cases}). \end{cases} \tag{2.4}$$

The subsequent decay of neutral pions, $\pi^0 \to 2\gamma$, leads to the production of gamma rays and the potential observability of the source in the electromagnetic spectrum, while neutrinos are produced in the subsequent decay of the charged pions in the same way as in the pp-scenario.

Pion decay produces neutrinos via:

$$\begin{aligned} \pi^\pm &\to \mu^\pm + \overset{(-)}{\nu}_\mu, \\ \mu^\pm &\to e^\pm + \overset{(-)}{\nu}_\mu + \overset{(-)}{\nu}_e, \end{aligned} \tag{2.5}$$

where the neutrino ν_α is produced with the anti-lepton l_α^+, and the anti-neutrino $\bar{\nu}_\alpha$ is produced with the lepton l_α^-, conserving lepton number. Thus, the neutrino flavor ratio from pion production at source is $\nu_e : \nu_\mu : \nu_\tau = 1 : 2 : 0$, where we have dropped the $\nu/\bar{\nu}$ distinction. On average, all neutrinos produced in pion decay receive $\sim 5\%$ of the parent proton's energy.

In sources with strong radiation or magnetic fields, muons interact before decaying and lose a substantial amount of their initial energy in the interactions. Thus, the resulting particles from muon decays have lower energies and do not contribute to the high-energy neutrino signal. In this so-called **muon-damped** production mechanism, only high-energy ν_μ and $\bar{\nu}_\mu$ are produced, leading to a source flavor ratio of $\nu_e : \nu_\mu : \nu_\tau = 0 : 1 : 0$. The probability for muons to interact increases with their energy, thus a gradual transition from $1 : 2 : 0 \to 0 : 1 : 0$ has been proposed [66]. Identifying such a

transition may be possible with the proposed *IceCube-Gen2* facility but is out of reach in the work presented here. A muon-damped source typically is also a **muon-beamed** source at lower energies, where the muons pile up [67]. At the energies where muon decay completely dominates the neutrino production, the neutrinos are produced with a flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 0$.
In pγ-sources with very high magnetic field, also the pions can lose energy via radiative losses, and effectively be removed from the neutrino production. If the source is optically thin to neutrons, neutrinos are mainly produced in the decays of high-energy neutrons in the so-called **neutron beam** scenario [68] via

$$n \to p + e^- + \bar{\nu}_e. \tag{2.6}$$

Thus, only $\bar{\nu}_e$ are produced at source, leading to $\nu_e : \nu_\mu : \nu_\tau = 1 : 0 : 0$ and $\nu : \bar{\nu} = 0 : 1$. While the neutrino-nucleon cross-sections for anti-neutrinos are almost the same as for neutrinos at the energies at which the IceCube signal is dominated by the astrophysical neutrino component, the Glashow resonance (GR) [69] at $E_\nu \approx 6.3$ PeV can be used to probe the amount of $\bar{\nu}_e$ reaching IceCube. As neutrinos from neutron decay only carry $\sim 0.1\%$ of the parent neutron's energy, neutron decay only contributes to the observable neutrino flux if neutrino production from pion decay is inhibited. The parent neutrons can be produced via the Δ-resonance in Equation 2.4 or from the photodisintegration of heavy nuclei. A previous measurement of the flavor ratio disfavors the neutron decay scenario at 3.6σ [32].
Finally, the heavy mesons produced in a pγ-source may dominate the neutrino production if pions interact before decaying. This **charm-production** scenario [70] is equivalent to the prompt component of atmospheric neutrinos (see Section 2.6). The decay of charmed D and Λ_C mesons leads to the production of equal numbers of electron- and muon-neutrinos. Rare decays of $D_s, D^{0,\pm}$ can also produce ν_τ, however, the ν_τ component always stays at least an order of magnitude below the ν_e and ν_μ components [46]. The charm-production scenario leads to a source flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 :\leq 0.1$. In general, ν_τ production at sources is negligible in any conceivable scenario.
At ultrahigh energies (UHE), **cosmogenic neutrinos** can be produced from the decay of π^+ created when ultrahigh-energy protons interact with photons from the (CMB). This interaction would limit the maximum energy for CR protons, known as the **GZK-cutoff** for UHECR protons [40, 41], and produce GZK neutrinos. GZK neutrinos have not been observed yet. The decay of π^0 also created in those pγ_{CMB}-interactions would produce a diffuse gamma-ray flux. *Fermi-LAT* measurements can be translated into an upper limit

for cosmogenic neutrinos of $E_\nu^2 \Phi_\nu \leq 10^{-8}\text{GeVcm}^{-2}\text{sr}^{-1}\text{s}^{-1}$ at neutrino energies ~ 1 EeV [71]. No neutrinos have been observed in this energy range yet. Currently, the most stringent upper limits are obtained with IceCube's all-flavor Extremely-High-Energy (EHE) event selection and constrain the total neutrino flux to $E_\nu^2 \Phi_\nu < 2 \cdot 10^{-8}\text{GeVcm}^{-2}\text{sr}^{-1}\text{s}^{-1}$ at neutrino energies of 1 EeV [72]. The interpretation of anomalous events reported by the ANITA collaboration as UHE neutrinos is in strong tension with limits from IceCube and the Pierre Auger observatory, see e.g. [73].

2.3 Neutrino Propagation

Neutrino oscillation describes the phenomenon of neutrinos changing flavor during propagation. The concept was not accepted yet when the first measurements of solar neutrinos were conducted at the Homestake mine. Comparing predicted neutrino interaction count rates based on calculations of the solar neutrino flux and the detector efficiency [11, 74] to the observed count rate, a deficit was observed [12]. This lead to the so-called solar neutrino problem, which could only be solved decades later by the discovery of neutrino oscillations [75, 76, 77], and their resonant enhancement in matter [78, 79]. This explained the lower amount of neutrino interactions at the Homestake experiment which was only sensitive to ν_e. Neutrinos were long thought to be massless, and in the standard model (SM) of particle physics they are described as massless. However, neutrino oscillations can only be explained by mass differences between the three neutrinos, immediately implying at least two non-zero neutrino masses, as well as a rotation of the flavor eigenstates α wrt. the mass eigenstates i. So far, only upper limits on the absolute neutrino mass scale exist. A recent combination of cosmological measurements yields an upper limit to the sum of neutrino masses of $\sum_i m_i < 0.12$ eV [4]. Very recently, the limits on the effective electron neutrino mass in beta decay were improved for the first time in more than two decades, and are now $m_\nu < 1.1$ eV [5]. These masses do not have the exact same meaning, the subtleties however are not important for the analysis presented. The cosmological measurements are more constraining, but are model-dependent, while the measurements performed using beta-decay electrons have very weak model dependencies, but are not as constraining. A third measurement could be obtained using neutrino-less double-beta decay, but only if this process is allowed.

2.3.1 Neutrino Masses

The origin of neutrino masses is one of the mysteries of particle physics. In the SM, there are no right-handed neutrinos, and therefore neutrinos are massless. To obtain

mass terms, one has to introduce right-handed neutrino fields, which do not participate in weak interactions. If neutrinos are Dirac particles, like quarks and charged leptons, their mass can be generated by the Higgs mechanism, too. This leads to the Dirac mass term in the Lagrangian

$$\mathcal{L}^D = -\bar{\nu}'_R \mathcal{M}^D \nu'_L + h.c \tag{2.7}$$

where ν'_R are the right-handed neutrino fields. The mass terms then arise following the diagonalization of the matrix $\mathcal{M}^D$,

$$\mathcal{L}^D = -\sum_i m_i \bar{\nu}_i \nu_i. \tag{2.8}$$

This mechanism does not however explain the smallness of the neutrino masses, requiring Yukawa couplings orders of magnitude smaller for neutrinos than for quarks and charged leptons.
Another possibility arises if neutrinos are Majorana particles, i.e. their own antiparticles. The Majorana Lagrangian reads

$$\mathcal{L}^M = -1/2(\bar{\nu}'_L)^c \mathcal{M}^M \nu'_L + h.c., \tag{2.9}$$

and following the diagonalization of $\mathcal{M}^M$ one obtains the mass terms

$$\mathcal{L}^M = -1/2 \sum_i m_i \bar{\nu}_i \nu_i. \tag{2.10}$$

To generate the small neutrino masses, the seesaw mechanism [80, 81, 82] is employed. As right-handed neutrinos do not participate in weak interactions, their mass is not constrained to the electroweak scale. In the seesaw mechanism, the right-handed neutrino mass term is very high and naturally suppresses the resulting neutrino mass. The Majorana nature of neutrinos would imply a violation of the total lepton number conservation, which would allow neutrinoless double-beta decay to happen.

2.3.2 Neutrino Oscillations

First devised in 1957 by Bruno Pontecorvo as neutrino–anti-neutrino oscillations [13], the idea was later developed into the three-flavor neutrino oscillation framework known today, in which neutrino change flavor, but the initially studied neutrino–anti-neutrino oscillations do not occur [14]. Neutrino flavor eigenstates α are related to the mass

eigenstates i via a rotation matrix:

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i}|\nu_i\rangle. \tag{2.11}$$

In the standard three-flavor framework $U_{\alpha i}$ is the 3×3 Pontecorvo–Maki–Nakagawa–Sakata (PMNS) mixing matrix [13, 14] containing three mixing angles θ_{ij} and one CP-violating phase δ_{CP}, $\alpha = \{e, \mu, \tau\}$, $i, j = \{1, 2, 3\}$. If the masses of the neutrinos ν_i are different, their relative phases will change during propagation, leading to neutrino oscillations. The PMNS-matrix reads:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \cdot \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta} & 0 & c_{13} \end{pmatrix} \cdot \begin{pmatrix} c_{12} & 0 & s_{12} \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathcal{P}, \tag{2.12}$$

where the shorthand notation $s_{ij} \equiv \sin\theta_{ij}$ and $c_{ij} \equiv \cos\theta_{ij}$ is used; δ is the CP-violating phase; $\mathcal{P} = \mathbb{1}_3$ if neutrinos are Dirac particles and $\mathcal{P} = \mathrm{diag}(e^{i\alpha_1}, e^{i\alpha_2}, 0)$ with the additional Majorana phases $\alpha_{1,2}$ if neutrinos are Majorana particles. The oscillation probability reads:

$$P_{\nu_\alpha \to \nu_\beta} = |\langle \nu_\beta | \nu_\alpha(L) \rangle|^2 = \sum_{i,k} U^*_{\alpha i} U_{\beta i} U_{\alpha k} U^*_{\beta k} \mathrm{e}^{i(E_k - E_i)L} \tag{2.13}$$

with the baseline L. For relativistic neutrinos, we can approximate $E_k = \sqrt{p_k^2 + m_k^2} \simeq p_k + m_k^2/2p_k$. Setting all leading momenta p_k to be equal to a common energy E for all propagating neutrino mass eigenstates yields $E_k \simeq E + m_k^2/2E$. Substituting $E_i - E_k = \Delta m^2_{ik}/2E$ with $\Delta m^2_{ik} = m_i^2 - m_k^2$ one arrives at

$$P_{\nu_\alpha \to \nu_\beta} = \sum_i |U_{\alpha i}|^2 |U_{\beta i}|^2 + 2\mathrm{Re}\left(\sum_{i,k>i} U^*_{\alpha i} U^*_{\beta i} U_{\alpha k} U^*_{\beta k} \mathrm{e}^{i\frac{\Delta m^2_{ik} L}{2E}} \right). \tag{2.14}$$

The second term of Equation 2.14 is the oscillation term which depends on the neutrino energy E and the traveled distance L. Evidently, in order to measure neutrino mixing parameters, it is advantageous to have a detector with a good energy resolution placed at a known distance from a source, such that maximal dip in neutrino flavor disappearance (or peak in neutrino flavor appearance) is accessible by the experiment. As neutrino sources are not monochromatic in neutrino energy, and the baselines not well defined, the oscillatory term can be ignored for extragalactic neutrinos from several sources. We

then arrive at the average oscillation probability

$$\langle P_{\nu_\alpha \to \nu_\beta} \rangle = \sum_i |U_{\alpha i}|^2 |U_{\beta i}|^2. \tag{2.15}$$

Comparing Equation 2.15 with the values of U from Equation 2.12, one can see that the neutrino flavor composition that can be measured on Earth depends on all the mixing angles θ_{ij} as well as the CP-violating phase δ in addition to the initial flavor composition at the sources. By now, most of the neutrino mixing parameters have been measured by a range of dedicated experiments as well as with atmospheric neutrinos in IceCube. However, while the values of the Δm^2_{ij} have been measured, the sign of the mass-squared difference is only known for Δm^2_{12}. This leads to two possible mass orderings for neutrinos: In the **normal ordering** (NO), $m_1 < m_2 < m_3$, while in the **inverted ordering** (IO), $m_3 < m_1 < m_2$. In this thesis, we use the latest three-neutrino global-fit parameters [2] as provided by `NuFit4.1` [30, 83], which are summarized in Table 2.1. Note that the mass ordering and the $CP-$violating phase δ are still largely unconstrained, although data indicate towards $\delta \neq 0$ and the NO is preferred over IO in the latest fit. The mixing matrix U as given by `NuFit4.1` reads:

$$|U|_{\text{best fit}} = \begin{pmatrix} 0.82142745 & 0.55031308 & 0.14970791 \\ 0.28831143 & 0.59155205 & 0.75295597 \\ 0.49207059 & 0.5892552 & 0.64081576 \end{pmatrix}, \tag{2.16}$$

$$|U|_{3\sigma} = \begin{pmatrix} 0.797 \to 0.842 & 0.518 \to 0.585 & 0.143 \to 0.156 \\ 0.233 \to 0.495 & 0.448 \to 0.679 & 0.639 \to 0.783 \\ 0.287 \to 0.532 & 0.486 \to 0.706 & 0.604 \to 0.754 \end{pmatrix}. \tag{2.17}$$

2.3.3 Expected Neutrino Flavor Composition on Earth

Due to neutrino mixing and most of the astrophysical neutrino flux observed by IceCube being of extragalactic origin, the neutrino flavor composition on Earth will be different from the source flavor composition. At the energies considered for this work and with baselines exceeding the size of our galaxy, neutrino oscillations average out. However, different source flavor compositions will lead to different Earth flavor compositions, making it possible to constrain source production mechanisms by measuring the flavor composition on Earth.

[2] The default version excludes atmospheric neutrino data from SuperKamiokande.

Parameter	Normal Ordering (best fit) bfp $\pm1\sigma$	Normal Ordering (best fit) 3σ range	Inverted Ordering ($\Delta\chi^2 = 4.7$) bfp $\pm1\sigma$	Inverted Ordering ($\Delta\chi^2 = 4.7$) 3σ range
$\theta_{12}[^\circ]$	$33.82^{+0.78}_{-0.76}$	$31.61 \to 36.27$	$33.82^{+0.78}_{-0.76}$	$31.61 \to 36.27$
$\theta_{23}[^\circ$	$49.6^{+1.0}_{-1.2}$	$40.3 \to 52.4$	$49.8^{+1.0}_{-1.1}$	$40.6 \to 52.5$
$\theta_{13}[^\circ]$	$8.61^{+0.13}_{-0.13}$	$8.22 \to 8.99$	$8.65^{+0.13}_{-0.13}$	$8.27 \to 9.03$
$\delta_{CP}[^\circ]$	215^{+40}_{-29}	$125 \to 392$	284^{+27}_{-29}	$196 \to 360$

Table 2.1: Three-flavour oscillation parameters relevant to this work from fit to global data. The numbers in the 1st (2nd) column are obtained assuming NO (IO), i.e., relative to the respective local minimum. See www.nu-fit.org or [30] for all parameters.

Production scenario	Source composition	Best fit parameters $\nu_e : \nu_\mu : \nu_\tau$	3σ mixing parameter range ν_e	ν_μ	ν_τ
pion decay	1 : 2 : 0	0.30 : 0.36 : 0.34	0.29 – 0.34	0.34 – 0.37	0.32 – 0.34
muon-damping	0 : 1 : 0	0.17 : 0.45 : 0.37	0.16 – 0.24	0.40 – 0.48	0.36 – 0.38
neutron decay	1 : 0 : 0	0.55 : 0.17 : 0.28	0.52 – 0.58	0.16 – 0.24	0.22 – 0.30
charm production	1 : 1 : 0	0.36 : 0.31 : 0.33	0.35 – 0.39	0.31 – 0.32	0.29 – 0.33

Table 2.2: Resulting flavor scenarios on Earth for given source flavor scenario, and the neutrino mixing parameters.

In Table 2.1, the resulting flavor compositions on Earth are shown for the neutrino production scenarios and source flavor compositions discussed in Section 2.2 and the oscillation parameters given in Table 2.1.

Figure 2.3 depicts the resulting flavor compositions on Earth, obtained by varying the mixing parameters from Table 2.1 within their 3σ allowed ranges. `NuFit4.1` [30] provides the $\Delta\chi^2$ values for each fit parameter and for all parameters' two-dimensional covariances. The 3σ interval in four dimensions corresponds to an uncertainty region of $\Delta\chi^2 = 16.25$ in each dimension. For δ_{CP}, the values in $\Delta\chi^2$ do not reach that value, δ_{CP} is thus unconstrained. This results in a three-dimensional uncertainty problem such that the uncertainty region of interest is $\Delta\chi^2 = 14.16$ in each dimension and for each two-dimensional correlation. Higher-dimensional covariances are not provided, and are ignored. The resulting flavor compositions for the neutrino production mechanisms considered are shown in color. Further, the range of resulting flavor compositions for completely arbitrary source flavor compositions are shown in gray, with the region accessible only by allowing some ν_τ production at source in a lighter gray shade. The filled in region thus corresponds to the maximum allowed parameter space for the neutrino flavor composition to be measured on Earth, assuming only standard mixing. As can be seen, only a small region of the full triangle is allowed. Any measurement of a flavor

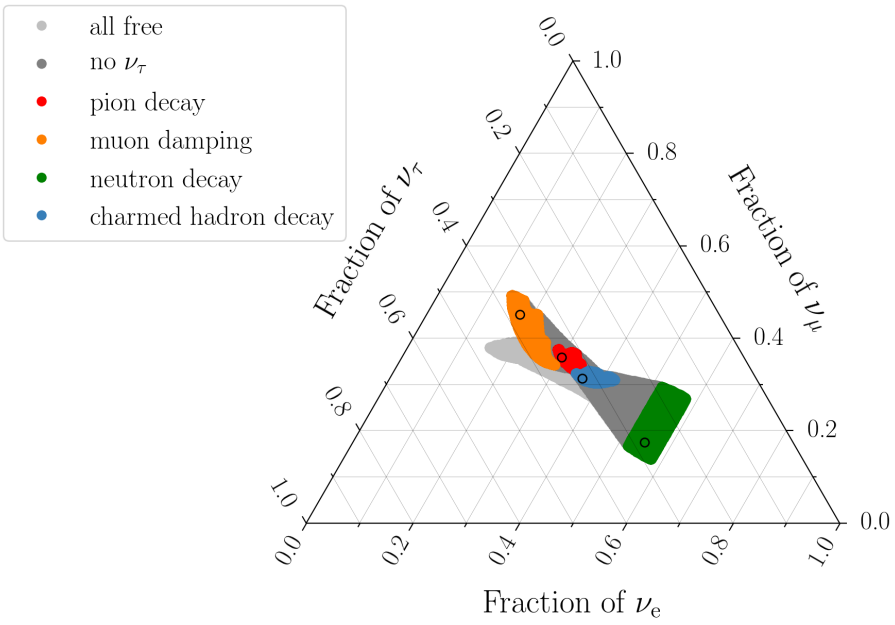

Figure 2.3: Range of possible resulting flavor composition on Earth assuming mixing parameters from [30]. The filled-in region shows the total allowed parameter space if the mixing parameters are free to vary within their 3σ uncertainties, and the source flavor composition is allowed to vary, with the light-gray region showing the part of the parameter space only accessible by a non-zero ν_τ fraction at the source.

composition outside of the filled-in region would point to non-standard neutrino propagation, altered e.g. by the existence of additional, sterile neutrino states, or neutrino decay. The influence of new physics on the on-Earth flavor composition is discussed in [31].

2.4 High-Energy Neutrino Interactions

Neutrinos are electrically neutral and rarely interact. They are only charged under the weak force and can only be detected indirectly, via the nuclear transformations caused by neutrino emission or capture, or, at the energies relevant for IceCube, via the particles they create in an interaction. At the energies accessible by IceCube, neutrinos interact mainly via deep inelastic scattering (DIS) with a nucleon initially bound in an atomic nucleus in the ice. Neutrinos can interact via charged-current (CC) or neutral current (NC) interactions:

$$\nu_\alpha + N \to l_\alpha + X \qquad \text{(CC)} \tag{2.18}$$

$$\nu_\alpha + N \to \nu_\alpha + X \qquad \text{(NC)}. \tag{2.19}$$

The neutrino of flavor $\alpha = \{e, \mu, \tau\}$ interacts with a nucleon N bound in an atomic nucleus. The energy transferred in the process breaks apart the nucleus and leads to the creation of secondary hadrons. As these hadrons keep producing more particles as long as they have a sufficiently high energy, a high-multiplicity shower develops. The final state consists of this hadronic shower and either a neutrino (NC interaction) or a charged lepton (CC interaction) of the same flavor α. The Feynman diagrams for these interactions are shown in Figure 2.4. In a NC interaction, the neutrino scatters off a nucleon bound in a nucleus, transferring a fraction of its energy to the nucleon which creates a hadronic cascade consisting of charged and neutral hadrons. The neutrino leaves the detector following the interaction. Thus, in NC interactions, the neutrino flavor cannot be identified. In a CC interaction, the neutrino creates a charged lepton of the same flavor in addition to transferring energy to the nucleus. In CC interactions, the neutrino is destroyed, but its flavor can be identified by identifying the flavor of the final state charged lepton. The neutrino-electron cross-section can largely be neglected except for one particular case: for a $\bar{\nu}_e$ with an energy of $E_{\bar{\nu}_e} \approx M_W^2/2m_e = 6.3$ PeV impacting on an electron at rest, the cross-section for $W-$boson production is resonantly enhanced. The resonance is very narrow, and is known as the Glashow resonance [69]. After being predicted in 1960, and sought after since the beginning of data taking with IceCube, recently the first strong candidate for a GR interaction has been found [84]. The GR interaction currently is the only way to discriminate between neutrinos and anti-neutrinos in IceCube. As no GR-scale interaction was observed in the high-energy starting events (HESE) sample that this work is concerned with, we will predominantly use ν to refer to neutrinos and antineutrinos.

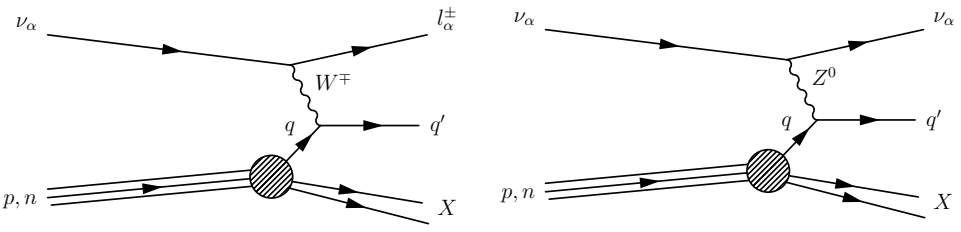

Figure 2.4: Feynman diagrams for deep inelastic scattering interactions of neutrinos. *Left:* Charged current interaction. *Right:* Neutral current interaction.

2.4.1 Lepton Energy Losses

To identify neutrino flavor, the **energy losses** of the charged lepton created in a CC-interaction need to be understood. The description is largely based on [39] and [38]. The energy losses are commonly given as $\mathrm{d}E/\mathrm{d}X$, the differential energy $\mathrm{d}E$ lost while traversing a differential amount of matter $\mathrm{d}X$. There are continuous losses due to ionization of the medium the leptons propagate through. In addition, with increasing energy, radiative energy loss processes become dominant over the ionization losses. Radiative losses are bremsstrahlung, pair production, and photonuclear interactions. Bremsstrahlung is caused by the deflection of the propagating charged particle through Coulomb scattering off a target atom. Pair production is a secondary loss, in which an emitted photon creates an electron-positron pair. Photonuclear interactions occur when an emitted photon disintegrates an atomic nucleus. The radiative losses are stochastic, and scale approximately linearly with the lepton's energy: $\langle \mathrm{d}E/\mathrm{d}X \rangle \simeq -E/X_0$. Here, X_0 is the radiation length, or average distance after which the lepton's energy has decreased to $1/e$ of its initial energy. As $X_0 \propto m^2$, the radiation length for each charged lepton is very different and consequently the pattern of the emitted radiation can be used to identify the propagating lepton. It is convenient to write $-\mathrm{d}E/\mathrm{d}X = a(E) + b(E)E$, where a are the ionization losses and b is the sum of all radiative losses.
Besides energy losses during propagation, a lepton may also decay if it is unstable. Electrons are stable particles and have a mass of $m_e = 511\mathrm{keV}/c^2$. Muons and taus, however, are not. The muon has a mass of $m_\mu = 105.7\mathrm{MeV}/c^2$, a lifetime of $\tau_\mu = 2.2 \cdot 10^{-6}$ s and decays via

$$\mu^- \to e^- + \bar{\nu}_e + \nu_\mu \tag{2.20}$$

and equivalently for μ^+. The tau has a mass of $m_\mu = 1.777\mathrm{GeV}/c^2$, a lifetime of $\tau_\tau = 290.3 \cdot 10^{-15}$ s and decays via

$$\tau^- \to \begin{cases} e^- + \bar{\nu}_e + \nu_\tau & (17.8\%) \\ \mu^- + \bar{\nu}_\mu + \nu_\tau & (17.4\%) \\ X + \nu_\tau & (64.8\%) \end{cases} \tag{2.21}$$

and equivalently for τ^+, where X denotes a hadronic cascade[3]. The probability for an unstable particle to propagate a distance l before decaying, is $P(l) \propto \exp(-l/L)$, where L is the particle's decay length. It is given by $L = \beta\gamma c\tau$ with the Lorentz factor $\gamma = 1/\sqrt{1-\beta^2}$, velocity β (in units of the speed of light, c), and lifetime τ of the par-

[3] The most common hadronic decay modes of the tau are $\tau^- \to \pi^- + \pi^0 + \nu_\tau$ (25.5%), $\tau^- \to \pi^- + \nu_\tau$ (10.8%), $\tau^- \to \pi^- + 2\pi^0\nu_\tau$ (9.3%), $\tau^- \to 2\pi^-\pi^+ + \nu_\tau$ (9.0%) and equivalently for τ^+ [38].

ticle. Using $E = \gamma mc^2$, we can write $L = \beta \frac{E}{mc} \tau$.

For **electrons**, we only care about the radiative losses, as electrons are stable. Their small mass also leads to a small radiation length: $X_0 = 39.31$ cm [38]. Thus, electrons radiate their energy away very promptly, the dominant energy loss process being bremsstrahlung, which leads to an electromagnetic cascade. At very high energies > 10 PeV, the Landau-Pomeranchuk-Migdal (LPM) effect [85, 86] leads to a suppression of bremsstrahlung, and therefore to an elongated cascade.

Neglecting radiative losses, a 1 TeV **muon** has a decay length of $L_\mu(E_\mu = 1 \text{ TeV}) = 6.2$ km. As $X_0 \propto m^2$, the muon radiation length of $X_0 \sim 17$ km is much larger than for the electron. Above ~ 1 TeV muon energy, the ionization losses scale only logarithmically with energy, while the radiative losses scale approximately linearly with energy. As $a(E = 1 \text{ TeV}) \sim 2$ MeV g^{-1} cm^2, and $b(E = 1 \text{ TeV}) = 3.17 \cdot 10^{-6}$ g^{-1} cm^2 [87], the radiative losses dominate over the stochastic ones in the energy loss term $-\mathrm{d}E/\mathrm{d}X = a(E) + b(E)E$. Thus, a high energy muon will traverse the detector without decaying, depositing energy losses due to Bremsstrahlung, pair production, and hadronuclear reactions along its path, with the average energy losses being proportional to the muon energy. This in principle enables the measurement of the muon energy, which however is complicated by the radiative losses being stochastic in nature. Thus, a muon can also "sneak" through the outer detector layers unnoticed and mimic a starting ν_μ−CC interaction. As the probability of depositing light is energy-dependent, higher-energy muons are brighter and more likely to deposit light as they enter the detector. Muon bundles contain a number of muons, which all have stochastic losses. These sum to an almost smooth loss profile, making muon bundles less problematic to identify as they enter the detector.

The **tau** has an even larger radiation length than a muon, $X_0 \sim 4754$ km, but its decay length is much shorter, $L_\tau(E_\tau = 1 \text{ TeV}) = 4.9$ cm. Thus, most taus created in IceCube will decay very promptly. 17.4% of the tau decays will produce a muon, which will then traverse the ice. The majority of 82.6% of tau decays will create hadrons or an electron, leading to a decay cascade. The hadronic cascade is darker than the electromagnetic cascade described above, this is due to the production of neutral particles or particles below the Cherenkov threshold. From the large radiation length as compared to muons, it is obvious that a tau will create much less light during propagation through the ice than a muon of the same energy. As will become clear from the next chapter, a 1 TeV tau cannot be distinguished from an electron due to its short decay length and the large sensor spacing of IceCube. However, the propagation of the tau over larger distances leads to the second decay cascade being resolvable in space and time with IceCube.

The event signature then consists of two causally connected cascades, called "double cascade", and is a central aspect of the analysis presented in this thesis. It is useful to write the decay length as $\langle L_\tau(E_\tau)\rangle \approx 50 \text{ m } E \text{ PeV}^{-1}$.

2.5 Diffuse Astrophysical Neutrino Flux

IceCube has reported the measurement of an astrophysical neutrino component, first observed with the **high-energy starting event** (HESE) selection [23, 88]. By now, the astrophysical component has been measured also in other channels, which we briefly describe below.

- **HESE** is a veto-based, all-flavor and all-sky selection of high-energy events that start within the detector volume. The fit is performed for reconstructed deposited electromagnetic-equivalent energy $E_{\text{dep}} \geq 60$ TeV. The astrophysical component stands out in the energy and zenith distribution wrt. the atmospheric backgrounds. Due to the veto region the effective volume is reduced. The latest results are based on ∼7.5 years of data. HESE is described in more detail in Section 4.2.

- The **through-going muons** (TGM) form a topology-based selection of up-going track-like events, containing mainly atmospheric and astrophysical ν_μ. The astrophysical component dominates starting at $E_{\text{dep}} \geq 200$ TeV. The effective volume is enlarged by also accepting events where the neutrinos interacted outside of the instrumented volume. The latest results contain ∼9.5 years of data, with the first year using the initial detector calibration (*pass-1*) [89].

- The **contained cascades** are a topology-based, all-sky selection of high-energy cascade-like events, with a small contribution from ν_μ-CC interactions. The astrophysical component stands out in the energy and zenith distribution wrt. the atmospheric backgrounds. The latest results contain ∼6 years of data, however, the first two years had a higher energy threshold than the later added years. This selection still uses the initial detector calibration (*pass-1*) [90].

A summary of the most recent measurements in these individual channels is given in Table 2.3 and in Figure 2.5. The astrophysical spectral index γ and normalization Φ are given for the assumption, that the spectrum forms a single power-law

$$\frac{\mathrm{d}\Phi_\nu}{\mathrm{d}E} = \Phi_{\text{astro}} \left(\frac{E_\nu}{100\text{TeV}}\right)^{-\gamma_{\text{astro}}} \cdot 10^{-18} \; [\text{GeV}^{-1}\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}]. \tag{2.22}$$

Analysis Sample	Livetime [Years]	Approx. E_ν[TeV]	Arrival Direction	Dominant Flavor	Φ_{astro}	γ_{astro}
HESE	7.5	50 – 5000	All-sky	e, μ, τ	$2.15^{+0.49}_{-0.53}$	$2.88^{+0.2}_{-0.19}$
Contained Cascades	6	5 – 5000	All-sky	e, τ	$1.66^{+0.25}_{-0.27}$	2.59 ± 0.07
TGM	9.5	50 – 10000	Northern Sky	μ	$1.44^{+0.25}_{-0.24}$	$2.28^{+0.08}_{-0.09}$

Table 2.3: Comparison of the single power-law fits performed on three diffuse astrophysical neutrino samples. γ_{astro} is the astrophysical spectral index, Φ_{astro} is the astrophysical normalization at $E = 100$ TeV with units 10^{-18} [$\text{GeV}^{-1}\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}$].

Note that the energy ranges are not the same for all three analyses. All analyses agree in their common energy range, however, the statistics get sparse for $E_{\text{dep}} > 200$ TeV for HESE and contained cascades.
In addition to these event selections, other event selections and measurements exist that are sensitive to the astrophysical component:

- The **partially contained cascades** is an extension of the contained cascades, focusing on events with vertices at the boundary of the detector. It has an energy threshold of $E_{\text{dep}} = 34$ TeV. This event selection has been applied to two years of data, with the resulting fit comparable with the contained cascades. Also a combined cascade fit has been performed with two years of data, giving a spectral index of $\gamma_{\text{astro}} = 2.67^{+0.12}_{-0.13}$ and an astrophysical normalization of $\Phi_{\text{astro}} = 2.3^{+0.7}_{-0.6}$ [26].

- The **PeV energy partially-contained events** (PEPE) is also a selection for events with vertices outside the instrumented volume, however, it is most sensitive to events even further outside the detector boundary and at even higher energies, with a focus on the $\mathcal{O}(1-10)$ PeV range. Using this event selection, the first GR candidate event was found [92, 84].

2.6 Atmospheric Backgrounds

Two types of particles produced in CR interactions in the Earth's atmosphere can reach the detector: **atmospheric muons** and **atmospheric neutrinos**. The collisions of CR with nuclei in the atmosphere produce cascades of secondary particles, among them muons and neutrinos. Atmospheric muons are responsible for the vast majority of IceCube events and thus pose a major problem when looking for the much rarer neutrino

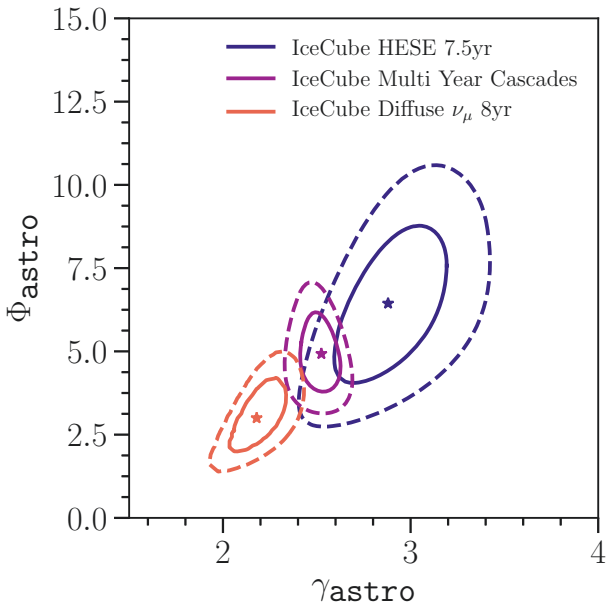

Figure 2.5: Single power-law comparison of the three diffuse analyses, adapted from [91]. The contours show the 1σ (full lines) and 2σ (dotted lines) confidence intervals. The measurements are in agreement within 2σ.

interactions. To reduce the muon contamination of a sample of neutrinos, characteristics in their flux, spectrum and created topology are employed. Duo to the muon's limited lifetime of 2.2 μs and their energy loss, only relativistic muons can reach the detector. On the way from their production site in the atmosphere to the detector, they lose energy which leads to a quicker decay. **Conventional atmospheric muons** are produced in the decay of pions and kaons created in CR interactions in the atmosphere. At energies relevant for this work, the conventional atmospheric muons follow a spectrum with $\sim E^{-3.7}$, steeper than the parent CR spectrum. At vertical incident angles, the muons have to traverse the shortest path through the atmosphere and the ice overburden to reach the detector. Muons with very horizontal incident angles come from showers at the horizon. They are in general more energetic, as they have to traverse more of the atmosphere and ice to reach the detector, leading to a decay of most of the lower energy muons. Furthermore, the warmer, less dense atmosphere at the horizon leads to the parent pions and kaons to interact less before decaying, therefore transferring more of their initial energy to the resulting muon. There are no upgoing atmospheric muons, i.e. muons traveling from deep in the Earth toward the surface, atmospheric muons, as

they decay en-route to the detector through the bedrock. However, a contamination of a sample from downgoing muons reaching IceCube from the Antarctic surface, which have been mis-reconstructed as upgoing, is possible. Most conventional atmospheric muons are not created by themselves, but in bundles of up to thousands of muons created in the same shower. At high energies, the prompt decay of heavy charmed mesons (mainly $D_s, D^{0,\pm}, \Lambda_C$) and unflavored mesons (η, ω, ϕ) creates **prompt atmospheric muons**, which approximately follow the incident CR spectrum.

Atmospheric neutrinos can also be subdivided into two components with different production mechanisms, spectral characteristics, and flavor compositions. **Conventional atmospheric neutrinos** are $\nu_{e,\mu}$ from the decay of $\pi^{\pm}$ and $K^{0,\pm}$ produced in the atmosphere in cosmic ray interactions. There is no conventional atmospheric ν_τ flux. Due to the interactions of secondary mesons in the atmosphere before decaying, conventional atmospheric neutrinos follow a spectrum with a spectral index one unit softer than the incident CR spectrum and their flux is peaked at the horizon. The conventional ν_e flux is greatly dominated by the decay of $K^{\pm}, K^0_L$ and above $E_{\nu_e} \sim 100$ TeV also by K^0_S. The conventional ν_μ flux is dominated by $K^{\pm}$ decays above $E_{\nu_\mu} \sim 0.1$ TeV, with a sub-dominating contribution from $\pi^{\pm}$ decays [46]. The flavor composition of the conventional component is $\sim \nu_e : \nu_\mu : \nu_\tau = 1 : 20 : 0$, where at low energies not relevant for this thesis the oscillation of ν_μ to ν_τ leads to a modification of the flavor composition depending on the zenith angle, which can be measured with the DeepCore array in the ν_μ disappearance [93] and ν_τ appearance channels [94].

At higher energies, the atmospheric flux becomes dominated by the **prompt component**, stemming from decays of charmed hadrons, of which rare decays of $D_s, D^{0,\pm}$ can produce $\tau - \nu_\tau$ pairs. As shown by [46], tau neutrino prompt fluxes always stay at least an order of magnitude below the muon neutrino prompt fluxes. The prompt atmospheric neutrino flux has not been measured yet. As the lifetime of the heavier charmed mesons is shorter than their interaction time, the prompt neutrinos follow a spectrum with spectral index close to the incident CR spectrum, and the flux is constant in the cosine of the zenith angle. The prompt component has a flavor ratio of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 :\leq 0.1$.

In addition to having a different energy spectrum than astrophysical neutrinos, atmospheric neutrinos are often accompanied by atmospheric muons born in the same CR induced shower. These accompanying muons have a zenith-dependent probability of reaching the detector and depositing light there. This fact can be used to suppress many downgoing atmospheric neutrinos: in the HESE selection via the muons triggering the veto, which will be described in Section 4.2.2, and in the contained cascades selection via the track-like topology caused by the accompanying muons. The TGM

selection accepts upgoing muons traversing the detector, as they must stem from neutrino interactions outside of the instrumented volume since any atmospheric muon would decay long before reaching the detector on its way through the bedrock.

Chapter 3

The IceCube Detector

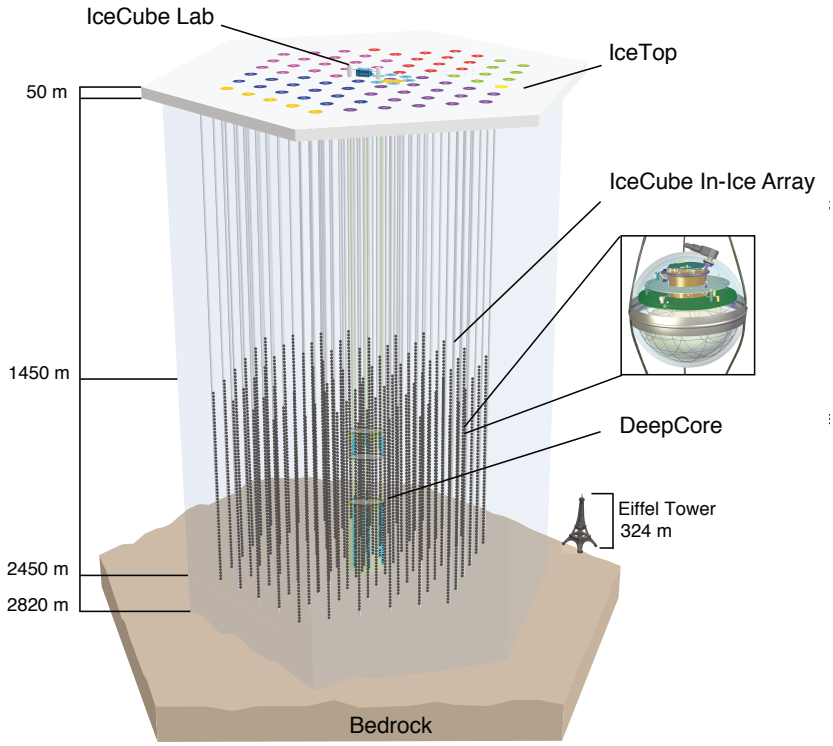

Figure 3.1: The IceCube detector, modified from [95]. The subarrays IceTop, DeepCore as well as the IceCube Lab are indicated. A Digital Optical Module is highlighted.

The IceCube South Pole Neutrino Observatory is a km^3 Cherenkov detector located at the Amundsen-Scott South Pole Station in Antarctica. It consists of 5160 Digital Optical Modules (DOMs) arranged on 86 strings at depths between 1450 m and 2450 m below the

surface. IceCube detects neutrinos with energies between ~ 5 GeV and $\mathcal{O}(10)$ PeV via the Cherenkov light induced by secondary charged particles as they pass through the ice. IceCube is a unique detector serving multiple purposes including astrophysical neutrino measurements and neutrino astronomy, atmospheric neutrino oscillations measurements, and searching for physics beyond the standard model. Currently, IceCube is the only experiment able to measure astrophysical neutrinos, but there are two large neutrino telescopes under construction in the Northern hemisphere. KM3NeT [96] is being built in the mediterranean sea, where the smaller prototype detector ANTARES [97] is still taking data. Due to its limited size, ANTARES does not have the sensitivity needed to measure the astrophysical neutrino component. The largest neutrino telescope in the Northern hemisphere is now Baikal-GVD [98], taking data while under construction in Lake Baikal in Russia. Once completed, both detectors will be comparable in size to IceCube.

The construction of IceCube took seven South Pole summers. Deploying > 5000 sensors in the antarctic ice to a depth of 2500 m is a challenging task. The top 50 m consist of compressed snow, called firn, which can be drilled through using a heated probe, the "firn drill". Below the firn layer, the ice was melted using hot pressurized water from a 5 MW enhanced hot water drill specifically designed for IceCube [99]. The holes were drilled to approximately 60 cm diameter, and 2500 m depth. The DOMs were attached to the in-ice cable shortly before being lowered into the hole. Once the holes are refrozen, the DOMs are immobilized and cannot be recovered. Before being shipped to the South Pole, each DOM was tested following a testing protocol to ensure reliable operation for at least 20 years in the ice. A final test was performed at the South Pole before deployment. At present, $> 98\%$ of the deployed DOMs are operating, with no signs of DOM aging [22]. The majority of the failed DOMs failed during deployment, some failed as a result of power outages, calibration runs, or firmware upgrades.

In Section 3.1 the main components of IceCube will be introduced, followed by a description of the data taking process in Section 3.2, which are described in great detail in [22]. The ice optical properties are described in Section 3.3. In Section 3.4 the detection of neutrino interactions can be in IceCube is described, the event topologies are introduced and event reconstructions are explained.

3.1 Detector Components

The design and construction of IceCube profited greatly from the experience gained in constructing and operating the AMANDA detector [21], which can be seen as a prototype

of IceCube. The mechanical design of the DOMs was heavily inspired by AMANDA's OMs, and the detector layout with DOMs on strings also drew from the AMANDA design [21].

What we will refer to as **IceCube** in the following is the IceCube in-ice array, the main part of the IceCube Neutrino Observatory as shown in Figure 3.1. This array consists of 78 strings arranged on an approximately hexagonal grid in x and y, with 60 DOMs on each string. The interstring spacing is ∼125 m and the vertical spacing between DOMs is 17 m. While the lower energy threshold is 100 GeV, the main array was designed to have the highest sensitivity to neutrinos in the $\mathcal{O}$(TeV) to $\mathcal{O}$(PeV) energy range. The analysis presented here only uses information from DOMs in the main IceCube array.

In the center of IceCube, there are eight **DeepCore** strings. Both the strings and the DOMs on the DeepCore strings are more closely packed than in the main IceCube array, with a horizontal spacing of ∼50 m and a vertical spacing of 7 m. The smaller inter-module distance lowers the energy threshold to ∼5 GeV. DeepCore is mostly used for measurements of the atmospheric neutrino oscillation parameters (see e.g. [94] for a recent measurement of atmospheric tau neutrino appearance), as well as for dark matter searches. The seven central standard IceCube strings surrounding the DeepCore strings are also used for DeepCore analyses. DeepCore DOMs are excluded from the high-energy analysis presented here.

The **IceTop** surface array consists of 162 ice-filled tanks with 2 DOMs inside each. Two tanks form a station, and the surface stations' geometry is very similar to the in-ice array's footprint. IceTop is used to measure cosmic rays with primary energies of $\mathcal{O}$(PeV) to $\mathcal{O}$(EeV). In the denser packed inner part of IceTop, the energy threshold is lowered to $\mathcal{O}$(100 GeV). IceTop is not used in any part of the analysis described here.

All of IceCube's DOMs are connected via twisted copper wire pairs, which provide power and communication, to the computers housed in the **IceCube Laboratory** (ICL) on the surface. The ICL houses the data acquisition system (DAQ). Here, data are filtered before being sent to the Northern Hemisphere via satellite link.

The Digital Optical Module

The DOM is IceCube's autonomous light sensor unit. Its components are shown in Figure 3.2. It consists of a 10" Hamamatsu photomultiplier (PMT) facing downwards [100], a main board containing the readout electronics, and calibration devices [101], which are all contained in a spherical glass housing, omitted in the figure. The glass housing provides the necessary protection of the components against the high pressure in the deep ice. An optical gel connects the PMT's photocathode to the glass, providing optical

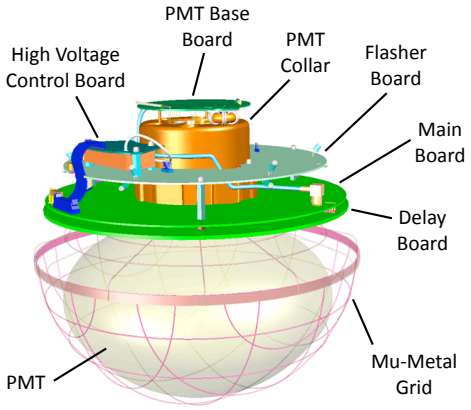

Figure 3.2: Components of the IceCube DOM. The glass housing and optical gel coupling between the PMT and the glass have been omitted. Figure taken from [22].

coupling and mechanical support. Further, a wire mesh surrounds the PMT bulb, acting as magnetic shield. The main board is responsible for the control of all DOM devices, communication to neighboring DOMs on the same string as well as the surface DAQ, digitization of the PMT waveforms as well as their compression and temporary storage, calibration of the internal DOM clock and the PMT gain. A flasher board is mounted on top of the mainboard and houses 12 Light-Emitting Diodes (LEDs) arranged in pairs with a 60° separation. In each pair, one LED emits light horizontally through the ice, while the other emits light upward at an angle of 48° wrt. the horizontal plane.
The PMT registers the signals created by photons induced by energetic charged particles propagating through the ice. Once a PMT registers a signal passing the discriminator threshold set to 0.25× the signal expected from a single photoelectron (PE), readout of the waveform is started. This waveform is digitized by the DOM's multiple digitizers. Their overlapping dynamic range and sampling speed enables the collection of a wide variation of signals. These range from single detected photons to high numbers of photons arriving simultaneously at a DOM from nearby, high energy cascades created in the interactions of PeV neutrinos. Analog Transient Waveform Digitizers (ATWD) record 128 samples of the waveform, spaced 3.3 ns apart, for a total duration of 427 ns, with three amplifier gains covering the PMT dynamic range. Signals arriving at the DOM up to 6.4 μs after the start of sampling are recorded on fast Analog to Digital Converters (fADC) at a resolution of 25 ns. This allows for waveform capture of delayed signals, which for example are produced by photons that have traveled from further away and undergone scattering. Each DOM can be uniquely identified via its ID, which it trans-

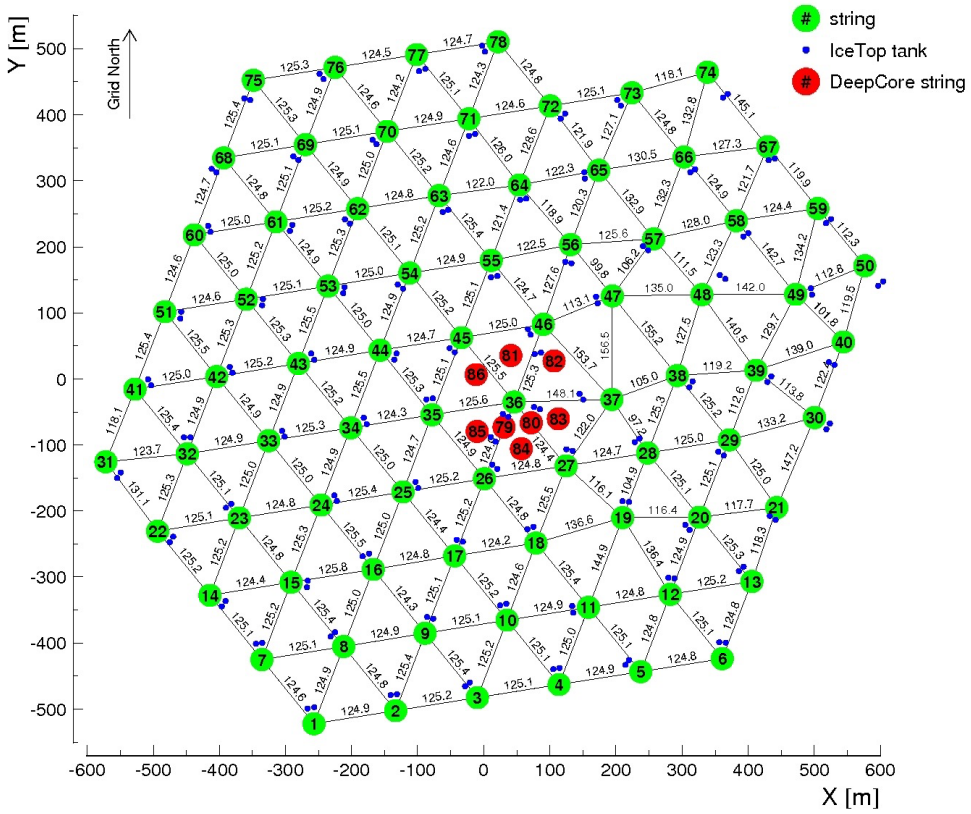

Figure 3.3: Top view of the IceCube detector, and the IceCube coordinate system. IceCube strings are shown in green, DeepCore strings in red, and IceTop tanks in blue. The origin of the IceCube coordinate system is close to the geometric center of IceCube.

mits to the surface along with the time-stamped waveform. The DOM ID corresponds to the DOM's position which is known to ∼1 m, and the relative timing resolution of the PMT pulses is ∼1 ns.

The IceCube coordinate system

The IceCube Coordinate system uses a right-handed $x, y, z-$coordinate system centered at the geometric center of the IceCube in-ice array. The origin is at a depth of 1950 m. The $z-$axis points towards the surface, the $y-$axis is aligned with the Prime Meridian, and the $x-$axis is at 90° clockwise from the $y-$axis. The direction of an event is given by the azimuth ϕ and zenith θ_z angles. The azimuth angle spans 360°, where an azimuth angle of $\phi = 0°$ is aligned with the positive $x-$axis. The zenith angle spans 180°, where a zenith angle of $\theta_z = 0°$ is anti-aligned with the $z-$axis, i.e. the particle is traveling

straight downwards. A schematic of the top view of the detector with the coordinate system is shown in Figure 3.3.

3.2 Triggering, Filtering and Calibration

The average detector uptime is highly stable and exceeds 99%, while the "clean" uptime has steadily increased over the years of operation to currently ∼98%. "Clean" uptime means that the full detector is taking good data to be used in analyses. A typical data-taking run has a duration of 8 hours. Most of the detector monitoring is done remotely, but to ensure a quick solution of problems as they arise, two *winterovers* are present at the South Pole year-round. Downtime is accumulated during transition periods between runs, power outages at the South Pole, and component failures. Certain runs cannot be used for analysis. These include calibration runs and flasher runs. During repair of failed components or sudden loss of communications with individual strings, runs with a partial detector configuration are carried out. Data from these runs may or may not be deemed of sufficient quality for a given analysis; the analysis presented here requires a full detector configuration.
The trigger rate is about 2.7 kHz, with the vast majority of triggers being due to atmospheric muons. In contrast, only about one astrophysical neutrino with a reconstructed energy above 60 TeV is recorded per month. This vast difference in the event rate requires sophisticated filtering and reconstruction algorithms, as even a tiny fraction of misreconstructed background muons could potentially spoil the extraction of the spectral parameters of astrophysical events.

3.2.1 Triggering and Online Filtering

Once the PMT signal exceeds the threshold corresponding to 0.25 PE, readout of the waveform is started, and data is collected for 6.4 μs. The noise rate is continuously monitored and is stable over time to $\sim 1\%$ [22]. Sources of noise include electrons emitted from the cathode of the PMT without a light source outside of the DOM being present, e.g. due to radioactive decays, scintillation in the glass of the PMT or the DOM pressure sphere, field emission within the PMT, or thermionic and electronic noise. A sudden, coherent increase in the DOMs' noise rate can indicate a Galactic supernova, with many neutrinos with energies in the MeV range interacting in IceCube simultaneously, each producing at most a tiny signal in a single DOM [102].
Each DOM is wired to its two nearest neighbors on either side, and if (next-to-)nearest neighbors of a DOM also register a photon within 1 μs, the full information on the de-

tected waveforms is digitized and these so-called "Hard Local Coincidence" (HLC) hits are transmitted to the IceCube Laboratory on the surface. Otherwise, the information is compressed and reduced in the DOM before being sent up. Requiring local coincidence greatly reduces the rate of noise-only triggers. The information arriving at the ICL consists of the time-stamped, digitized waveforms, and the DOMs' IDs. At the ICL, pulses are extracted from the digitized waveforms. Their collection enables reassembling the full event with high resolution in both time and space at any later point. Dedicated trigger algorithms, that search for causal connections between DOM hits, such as spatial or temporal patterns, are applied to the pulses. The fundamental trigger in IceCube is the Simple Multiplicity Trigger, which for in-ice DOMs requires eight HLC hits within 5 μs. The DAQ then builds the hits passing a trigger into events, and the online filter selects more interesting events whose data is sent to the Northern Hemisphere via satellite. To reduce the data stream for satellite transfer, reconstructions are performed on the raw events and multiple filters are applied. Due to computational limitations, only fast reconstructions are applied online. These are typically computationally inexpensive algorithms, which only use a limited amount of data and provide an initial estimate of an event's energy and direction to be used in the online filtering process. The filters are designed to select all potentially interesting events for subsequent processing, while removing obviously uninteresting events. One of these filters selects events with a high accumulated total number of PE, thus greatly reducing the amount of dim, lower-energy events. The data transmitted consists of the collection of digitized waveforms called a **pulsemap**, the passed filters, as well as the reconstructions applied online [22]. In addition, the detector **Geometry** containing the spatial information of the DOMs, the **Calibration** information of each DOM, and the **Detector** status are sent out in a file called GCD-file. The GCD-file contains all detector-specific information about which DOMs were taking data during the run, how well each DOM was operating, and all the DOMs' positions.
All processed events are stored locally on disk, and ∼15% of all triggered events are sent out via satellite after having been processed and compressed.

3.2.2 Offline Filtering

The next stages of filtering happen in the Northern Hemisphere. In each step, more sophisticated and computationally expensive reconstruction algorithms are applied to the data, and the amount of data is reduced further. The filters have been designed to select certain event types and reject events not belonging to that event class. One discrimination method is the event topology: track-like events that pass the Muon filter

at the South Pole get improved track reconstructions offline. Cascade-like events are selected by their approximately spherical shape through the Cascade filter at the South Pole, that rejects obvious track-like events. In the subsequent cascade processing, more sophisticated cascade reconstruction algorithms are applied, designed to filter out less obvious tracks while keeping all cascade-like events. Both these filters are used in high-energy event selections such as the Muon filters in the through-going muon [24] and the Cascade filters in the high-energy contained cascades [103] selections. The high-energy starting events selection uses the HESE filter, consisting of a simple veto and minimal number of PE cut, which in a one-step selection process reduces the data rate to a frequency of about one event per month. However, in the subsequent processing of the events, reconstruction algorithms from both the cascade and muon filtering chains are applied, as they provide a good seed for the final level processing. The HESE selection is described in Section 4.2, and the full HESE processing chain is described in Section 5.1.

3.2.3 Calibration

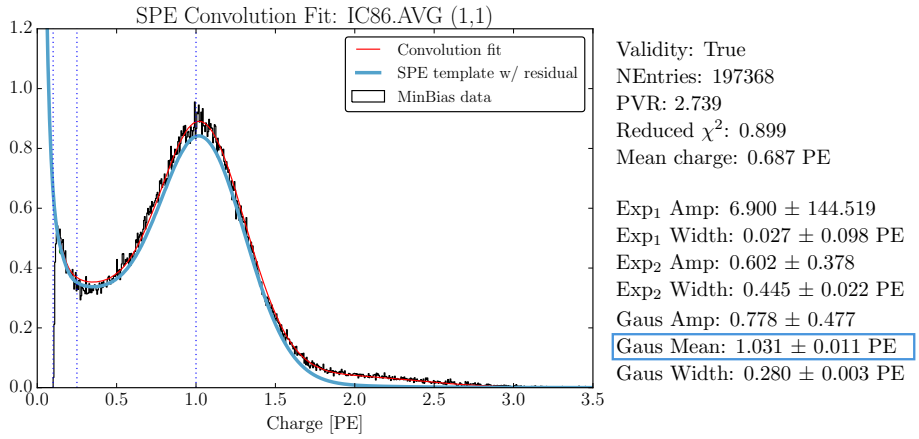

Figure 3.4: Charge distribution for DOM(1,1). The SPE charge template is shown in blue. It is extracted from a fit including the 2 PE contribution, shown in red. The fit parameters are given, with the SPE peak highlighted. It is fit to 1.031 ± 0.011 of the previously assumed value for this particular DOM. Figure reproduced from [104].

A precise calibration of all detector components is very important in order to translate the signal recorded by the DOM to the number of observed photons. This information is crucial in reconstructing the properties of particle cascades created by neutrino interactions. An important result of the calibration is the PMT's response to single

photoelectrons (SPE). The SPE peak is the peak of the charge distribution caused by a single photoelectron emitted from the PMT's photocathode. It is used to infer the total number of PE from the accumulated charge registered by the DOM. After the accumulation of several years of calibration data, a mismatch between the assumed and observed position of the SPE peak could be identified. An example fit is shown in Figure 3.4. This mismatch led to an overestimation of the energy of neutrino interactions by $\sim 5\%$ on average, well within the estimated uncertainties on the energy reconstruction of $\sim 15\%$ [105]. Following the new, improved calibration, all recorded data have been reprocessed, correcting that bias. The reprocessed data are called *pass-2* data, as opposed to *pass-1* data that use the initial, biased IceCube calibration. The latest HESE selection presented in this thesis uses *pass-2* data. The PMT waveforms are calibrated using a dedicated DOM calibration software [22].
The LED flashers deployed on each DOM are used to calibrate the DOMs' positions and response *in situ* and to study the optical properties of the ice. For this purpose, the signal received on the DOMs surrounding the emitting DOM is analyzed, as this response carries valuable information about the ice between emitter and receiver. Further, the performance of dedicated IceCube analysis software, e.g. the double cascade reconstruction algorithm described in more detail in Section 3.4.3, has been verified as the direction and origin of light emitted from the LED is known to a high precision [106]. Each LED can be flashed independently, but also a combination of multiple LEDs on one DOM flashing simultaneously can be configured. A simultaneous flashing of LEDs on different DOMs is not possible. The LED flasher runs are used to create ice models that contain the best knowledge of the ice optical properties.
In addition, mimimum-ionizing muons (MIMs) can be used for calibration. They have a nearly constant energy deposition along their path in the ice. The advantages of using MIMs are that they produce Cherenkov light and that their continuous energy loss is well known, while the disadvantages are the MIMs' unknown paths through the ice and the difficulty in obtaining a pure sample of MIMs.

3.3 Optical Properties of the South Pole Ice

Using the built-in calibration devices, as well as data taken during the construction phase of IceCube, the optical properties of the South Pole ice can be investigated in great detail, and ice models built for use in simulation and reconstruction. The untouched part of the ice is referred to as **bulk ice**. It consists of ice sheets that have formed over hundreds of thousands of years. In shallow ice, air bubbles are present in the ice, inhibiting

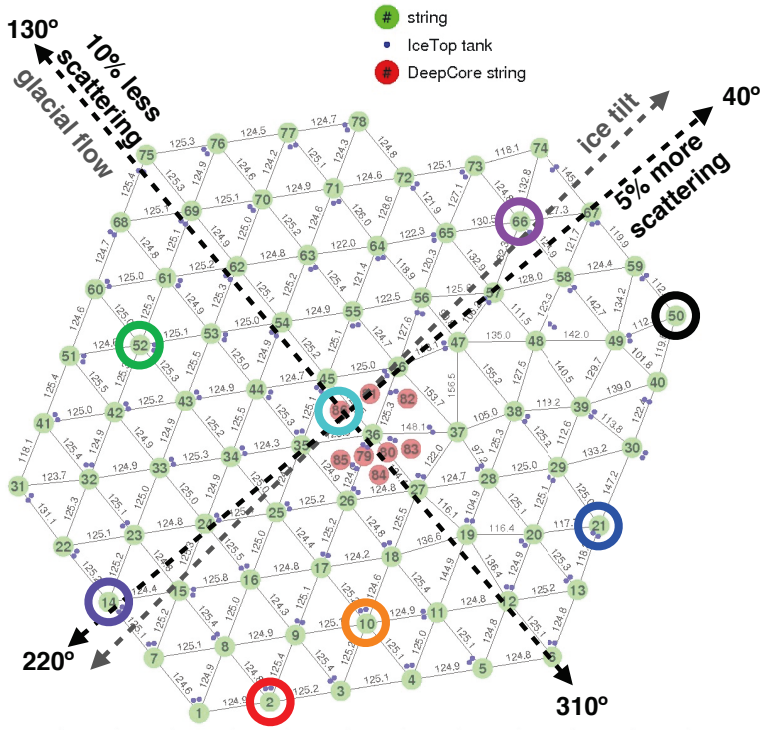

Figure 3.5: Top view of the IceCube with the anisotropy axes, the ice tilt axis and holes with dust logger data indicated. The major anisotropy axis is along the glacial flow (130° in the IceCube coordinate system). The minor anisotropy axes are orthogonal to the glacial flow, in the $x-y-$plane (40°) and along the positive $z-$axis. The ice layers are tilted, with the tilt axis approximately perpendicular to the glacial flow. The colored circles indicate boreholes from which dust concentration data was collected.

light propagation. With increasing depth and rising pressure, the air transitions from a gaseous to a solid air-hydrate clathrate phase. Below a depth of 1500 m where IceCube begins, the transition is complete and the ice is optically very transparent [107]. The ice sheets are tilted wrt. the horizontal plane in the IceCube coordinate system, with the major tilt axis being approximately perpendicular to the glacial flow as shown in Figure 3.5. As shown in Figure 3.6, the tilt changes as a function of the depth, and is largest closest to the bedrock where the influence of the bedrock's irregular shape is largest. Further, the scattering of photons is anisotropic, i.e. there is a preferred direction in which photons propagate. The melted and refrozen ice in the boreholes is referred to as **hole ice**. Hole ice has different optical properties than bulk ice, especially in the

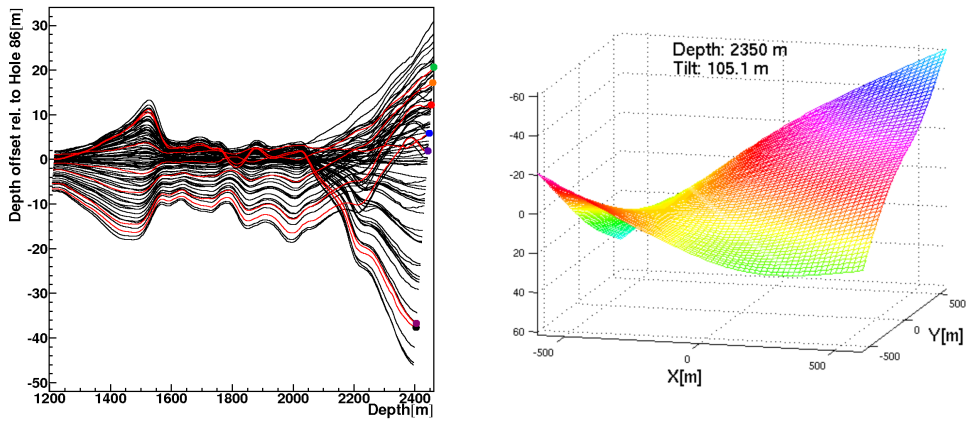

Figure 3.6: The ice layer tilt as obtained from the dust concentration measurements, taken from [108]. The ice sheets' offset is given relative to the most central borehole with the boreholes from which data was taken highlighted in red and a colored dot corresponding to the color of the rings in Figure 3.5 (*left*). The ice layer at the bottom of the detector is shown with the tilt relative to the central borehole indicated by color (*right*).

central region of the borehole containing air bubbles that became trapped during the hole re-freezing process. Two cameras have been deployed to monitor the refreezing of the hole ice and it's optical properties. However, due to the limited monitoring devices, the properties remain poorly understood. The South Pole Ice model applied to produce the simulation used in the analysis presented here is called **Spice3.2** [109]. The most critical parameters of light propagation through the bulk ice that are described by the ice models are:

- **Scattering and absorption:** The scattering length l_s is the average distance after which a photon will scatter during propagation. Each scatter leads to a change in the propagation direction of the photon, washing out the directional information of the incoming neutrino. In practice, only the effective scattering length $l_{\text{eff}} = l_s/(1 - \langle\cos\theta\rangle)$ can be measured, where $\langle\cos\theta\rangle$ is the average deflection angle at each scatter. The absorption length l_a is the distance after which the photon's survival probability drops to $1/e$, and therefore describes how far information from an event can travel. Scattering and absorption parameters are averaged over 10 m thick layers of ice and given for a wavelength of 400 nm [110]. Their wavelength dependence is known from using light sources emitting at different wavelengths [107].

- **Tilt:** The tilt parameter describes by how much the ice layers at each point in

space are tilted wrt. the center of IceCube. If the ice layer tilt is not accounted for, a wrong layer with different optical properties is used. As the ice optical properties are used to translate between light observed at the sensors and energy emitted in the ice, this can lead to large uncertainties and biases in the energy estimation and the angular reconstruction. These effects are particularly large for cascades in the deep ice and at the boundary of the detector. In the ice models, the tilt is modeled as having a maximum along the major tilt axis, and is set to zero perpendicular to it, which is an effective approximation found to sufficiently capture the effect of the tilt.

- **Anisotropy:** The propagation of photons through the ice at the South Pole is not isotropic [111], the cause of which is currently unknown. There is a preferred direction of propagation, which can be described to first order by a modulation of the scattering coefficients. The main anisotropy axis is in the x-y-plane, at an azimuth angle of 130°, and is described by 10% less scattering. Perpendicular to the main anisotropy axis, scattering is enhanced by 5% along an azimuth of 40° and in the zenith direction along the positive $z-$axis. The main ice anisotropy axis is thus aligned with the glacial flow. The effect of the anisotropy on the relative number of photons registered by DOMs 125 m away from the emitting LED is shown in Figure 3.7. The cause of the anisotropic light propagation is currently being investigated, recently e.g. the influence of the micro-structure of the ice polycrystals on the light diffusion has been calculated [112]. The anisotropy is an important systematic uncertainty to consider for ν_τ identification, as discussed in Section 5.4.

During the construction phase, so called "dust loggers" were lowered into several holes spread across the detector. Using their data, a dust concentration map was created [113, 108]. The dust concentration is constant per layer, and varies with depth. The highest concentration of dust is in depths between $\sim 2000 - 2100$ m and commonly referred to as the "dust layer". As photons scatter off of dust particles or get absorbed by them, the scattering and absorption coefficients are highest in the dust layer. Due to the drastically reduced photon propagation distance the optical properties of the ice have large uncertainties in the dust layer. It is also difficult to detect neutrino interactions in the dust layer and reconstruct their properties. The ice is most transparent below the dust layer, where l_s can reach up to 90 m, and l_a can reach almost 300 m for a wavelength of 400 nm. The ice layers are defined as planes of constant dust concentration. Comparing the depths of the layers as obtained from the dust logger data across the holes, it was found that the layers' depth varies. Interpolating between the holes

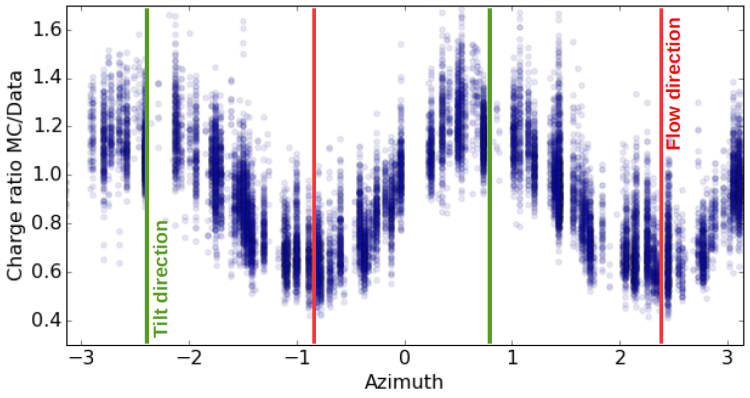

Figure 3.7: Optical ice anisotropy seen as azimuth dependent charge variations in flasher data. Along the glacial flow, more photons arrive at the DOMs than would be expected in isotropic ice, while along the tilt axis, less photons are registered. Figure taken from [112].

resulted in the construction of a detailed tilt map shown in Figure 3.6, which was used to derive the simpler parametrization with only one tilt axis that is implemented in the ice models.

The ice models have become increasingly more precise and complicated over time: Spice1 and Spice2 assumed homogeneous and isotropic ice, parametrized per 10 m layer. SpiceMie introduced the ice layer tilt and its fit was based on Mie theory of scattering [114], SpiceLea introduced the ice anisotropy [111]. Spice3 offered greatly enhanced statistics and therefore a better estimation of the uncertainties on the ice model parameters [109]. The anisotropy is still a parameter that is heavily studied, e.g. newest data seem to suggest that the anisotropy is depth dependent and might even vanish in the deep ice. It further seems to be the case that photons are preferably scattered in a way to align with the major anisotropy axis [112].

The hole ice mainly affects a DOM's angular response. As the water in the boreholes refroze, air bubbles formed and drifted towards the hole center. The diameter of the bubble column can affect the azimuthal response of a DOM, as the shorter the path the photon has to travel through the air bubble region, the higher the probability for that photon to reach the DOM's PMT. Currently, there are several models describing photons traveling through hole ice, whose parameters are not well constrained.

As the ice models are derived from data collected by the deployed DOMs, the optical properties are best constrained where the density of light sensors is high. Beyond the instrumented volume, the ice optical properties can only be extrapolated, leading to large systematic uncertainties on the reconstruction of events with vertices close to or

beyond the detector boundary.

3.4 Detection and Reconstruction of Events in IceCube

3.4.1 Cherenkov Detection

IceCube uses the Cherenkov effect [115] to detect relativistic, charged secondary particles created in neutrino interactions. When a charged particle travels through a medium, the moving charge will polarize the surrounding medium. The medium's electrons will be slightly displaced within the medium's structure. Moving back, they emit a photon. Once the charged particle exceeds the speed of light v_{phase} in the medium, the photons will interfere constructively in the Cherenkov cone with opening angle given by:

$$\cos\theta_C = \frac{1}{n\beta} \tag{3.1}$$

where θ_C is the Cherenkov angle, $n = c/v_{\text{phase}}$ is the refractive index of the medium, and $\beta = v/c$ is the particle's velocity. The condition $\beta > 1/n$, or equivalently $v > v_{\text{phase}}$ leads to a minimum energy at which particles create Cherenkov light, which in the ice with $n = 1.31$ is $E_{\text{min}} = 0.3$ MeV for electrons. For highly relativistic particles, $\beta \to 1$, and for IceCube the characteristic Cherenkov angle $\theta_C = 40.2°$ is obtained.
In a neutrino interaction, a hadronic shower is created, as described in Section 2.4. Many different particles can be created in the hadronic shower, making their precise calculation complicated. As only charged and highly relativistic particles produce Cherenkov light, and also neutral particles and particles below the Cherenkov threshold are produced in hadronic showers, the distinction between *total deposited energy* and *visible energy* of the shower must be made. The visible energy constitutes the energy transformed into light and coincides with the total deposited energy for electromagnetic showers. For hadronic showers, the relative light yield can be effectively described as [110]

$$f = 1 - (E_0/0.399\text{GeV})^{-0.130}(1 - 0.467) \tag{3.2}$$

in the energy ranges relevant for this analysis. It is $\sim 60\%$ at the lower energy range of IceCube and growing with energy, due to more π_0 being produced, decaying into photons [116].

3.4.2 Monte Carlo Simulation

Analyses in IceCube typically search for small signals buried under orders of magnitude larger noise. The analyses are kept *blind* during development, i.e. with no access to data. Thus, for each analysis the background rejection cuts are developed using Monte Carlo (MC) simulated events. They are then usually vetted on the "Burn sample", which typically contains $\sim 10\%$ of the data. In more recent analyses, it has also become common to verify the analysis on background regions to show data–MC agreement in parts of the parameter space not used for the analysis. Only after that, permission to *unblind* the full data set is granted by the collaboration. It has also become more common recently to look at the distribution of observables not used in the analysis in the signal region before using the unblinded data to determine the parameters of interest. Thus, a vast amount of MC events need to be simulated and processed.

Monte Carlo generation

The MC events are generated following a certain spectrum, for neutrinos typically a single power-law with a spectral index in the range of [1...2], and can be re-weighted to any spectrum for use in the analysis. Three software frameworks are used to generate different classes of MC events.

Neutrinos are generated using `NuGen`, an improved version of `ANIS` [117]. The neutrinos are generated in a volume around the Earth on paths that cross the detector. To avoid a waste of computational power, their interactions are forced to occur before they leave the volume from which light can reach the detector's DOMs. To account for forcing each neutrino's interaction, its calculated interaction probability enters its intrinsic weight. In IceCube, single atmospheric and astrophysical neutrinos can not be distinguished on an event-by-event basis. Thus, the `NuGen` generated neutrinos are used for atmospheric and astrophysical fluxes. Neutrinos will get weights assigned to them according to the expected (or tested) atmospheric and astrophysical fluxes, the weights can be used to calculate an expected event rate. Neutrinos and antineutrinos are usually generated in equal amounts, $\nu : \overline{\nu} = 1 : 1$, and each flavor is generated separately. Note that due to the tau's short lifetime and large radiation length, see Section 2.4.1, the tau typically transfers a large fraction of its energy to the tau neutrino produced in its decay. This leads to ν_τ regenerating and cascading down in energy during their propagation through the Earth, this effect is taken into account appropriately.

An adapted version of `CORSIKA` [118] is used to generate **muons** created in cosmic ray showers in the atmosphere. The generated primary particle is a nucleus, which then creates a particle shower in the atmosphere upon interaction. Of this shower, only neutrinos

and muons can reach the detector. However, as neutrinos rarely interact, the generation of atmospheric neutrino interactions using `CORSIKA` is impractical. Thus, only the muons created in the showers are propagated to the detector. The events can be weighted to different cosmic ray models. `CORSIKA` is currently the only software that produces muon bundles, i.e. large numbers of muons born in the same CR shower. `CORSIKA` propagates all particles born in a cosmic-ray shower, the majority of which is not detectable in a sub-surface detector like IceCube. This explains the major drawback of `CORSIKA`: it is computationally extremely expensive, as it is not targeted. Thus, to simulate single muons, `MuonGun` [119] was developed. It is much cheaper computationally, as single muons are generated directly around the detector volume. For HESE, which we will describe in Section 4.2, the main background indeed comes from single muons, as muon bundles deposit light more continuously and thus have a very high probability of being vetoed upon entering the detector.
To simulate random coincident events, muons from `CORSIKA` or `MuonGun` can be merged with the `NuGen` simulated neutrinos. An effect which is not currently simulated, is the accompanying of atmospheric down-going neutrinos by muons born in the same shower, which is very important for HESE. The probability of an accompanying muon to reach the detector and trigger the veto is instead calculated analytically; see Section 4.2.1 for details.
The generated MC events are **weighted**, i.e. they are drawn from a spectrum, which is suitable for the analysis, but does not necessarily reflect reality. A neutrino sample consists of n observed events over a livetime T, reflecting the rate $\mathrm{d}n/\mathrm{d}t$. The differential neutrino flux $\mathrm{d}\Phi/\mathrm{d}E$ can be written as:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}E} = \frac{\mathrm{d}n/\mathrm{d}t}{\mathrm{d}A\,\mathrm{d}\Omega\,\mathrm{d}E} \tag{3.3}$$

with unit time t, area A, solid angle Ω energy E. The weight of an event describes the ratio of the expected flux to the generated fluence,

$$w = \frac{\mathrm{d}n_{\mathrm{expected}}/\mathrm{d}t\,\mathrm{d}A\,\mathrm{d}\Omega\,\mathrm{d}E}{\mathrm{d}n_{\mathrm{generated}}/\mathrm{d}A\,\mathrm{d}\Omega\,\mathrm{d}E}. \tag{3.4}$$

Typically, the events are generated with a harder spectrum than realized in nature to provide enough MC events up to the highest energies relevant for an analysis. The events are re-weighted to an assumed or tested spectrum to obtain the expected rates at all relevant energies and directions.

Propagation of secondary leptons and photons

Once the generated neutrinos interact, the secondary leptons are propagated using the software `PROPOSAL` [120]. The electromagnetic and hadronic showers created by neutrino interactions in the ice are simulated [121] and the information about all particles and their energy losses stored in the event files.

After having created all secondary (and subsequent) particles and their energy losses, the resulting photons are propagated through the detector by `CLSIM` [122]. Photon propagation is usually the bottleneck of neutrino simulation, as it is computationally very expensive and currently can only be performed efficiently on GPU's, as all the photons' propagation through the ice need to be run in parallel. As the ice optical properties determine the scattering and absorption of photons in the ice, photon propagation needs an ice model. The events generated for the work described here use the Spice3.2 ice model. Typically, the photon light yield is artificially increased to 118% of the expected yield, this has proven useful for the following step.

Detector response

The final step of the event simulation is the **detector response**. Here, effects like the quantum efficiency of the PMT or the angular sensitivity of DOMs are taken into account and can be varied. The highest simulated efficiency of the DOMs is typically 118% of the nominal value, photons are down-sampled to obtain a set of simulated DOM efficiencies. Also, the knowledge of which DOMs are part of the detector configuration and taking good data is applied in this step. PMT effects like jitter, prepulses, or afterpulses, are added, as well as noise. At this step, also the same IceCube triggers and online reconstructions are applied as are applied to the data.

Afterwards, the simulation sets are processed in the same way as data, adding first the "Level 2" filters and reconstructions. Then at "Level 3", simulation is split according to topology into tracks (MuonL3) and cascades (CascadeL3). All-flavor analyses typically start at Level 2, and process simulation and data according to their needs, while topology-specific analyses usually start analysis-level processing from Level 3.

While a centralized `NuGen` and `CORSIKA` production exists, it has not been able to keep up with changes to the ice models, or produce high-energy sets with the same settings for each neutrino flavor. As HESE is an all-flavor event selection, and the work presented here is concerned with the flavor composition, all-flavor `NuGen` simulation was needed where the event generation, lepton and photon propagation, and detector response was performed in the exact same way for each flavor. Further, sets with the

same systematic variations also needed to be produced for all flavors. Therefore, dedicated simulation sets for HESE were produced[1]. It consists of baseline sets with nominal values for the ice optical properties and the DOM efficiency, as well as sets with discrete realizations of systematic variations for the bulk ice parameters like scattering and absorption, hole-ice parameters, and DOM efficiency. In addition, the simulation was produced for two distinct ice models, SpiceLea and Spice3.2. These sets cover an energy range of E_ν = [1 TeV...10 PeV] following an $E^{-1.5}$ spectrum. This choice of spectral index ensures that there are sufficiently large amounts of events at all energies relevant to the HESE event selection. This keeps statistical errors in the analyses small.

3.4.3 Event Structure and Reconstruction

The data collected by IceCube is stored as compressed `i3 files`, a data format developed for IceCube. Each i3 file is structured into frames. There are frames that describe the detector, namely the **Geometry**, the **Calibration** and the **Detector** status frames, which are independent of the data collected. Each event has a **DAQ** frame containing the raw data, such as the pulsemaps, and the **Physics** frame containing processing outputs, such as event reconstructions. All those frames are needed for the reconstruction of a single event: each reconstruction algorithm needs the data in form of a pulsemap, the knowledge of which DOMs to include, and their relative geometry as inputs. The full data $\vec{\eta}$ consists of the entirety of observed pulses: their charge (or average number of photoelectrons) k and their time t, on all DOMs $\{j\}$ with positions x_j, y_j, z_j. It reads $\vec{\eta} = (k_{jt}, x_j, y_j, z_j, t_{jt}, \Delta_{jt})$, where Δ_{jt} is the width of the time bin.
IceCube events are reconstructed using **maximum likelihood algorithms**. These algorithms maximize the likelihood of the data, i.e. the observed light pattern, for a given event hypothesis and are described in great detail in [105]. The likelihood to observe the data given the hypothesis is maximized by minimizing the negative log of the likelihood using gradient-descent minimizers. The number of photons detected on a DOM j is expected to follow a Poisson distribution with mean $\lambda_j = \Lambda_j E$, where Λ_j is the expected number of photons registered at DOM j coming from a 1 GeV cascade emitted at x, y, z, given the ice optical properties, and E is the cascade energy in units of GeV. The light yield Λ is determined using the IceCube flasher calibration, and is provided in tabulated form as photo spline tables [123].

[1] While the sets were produced specifically to be used for the new HESE analyses, they can be used by many high- or medium-energy selections. This is due to the fact that HESE is an all-sky, all-flavor and all-topology selection.

The likelihood for a cascade of energy E producing k photons on DOM j reads:

$$\mathcal{L}_j = \frac{(E\Lambda_j)^{k_j}}{k_j!} \mathrm{e}^{-E\Lambda_j}. \tag{3.5}$$

Using the logarithm and summing over all DOMs j, we get:

$$\ln \mathcal{L} = \sum_j k_j \ln(E\Lambda_j + \rho_j) - (E\Lambda_j + \rho_j) - \ln(k_j!) \tag{3.6}$$

where we also added the expected number of noise hits ρ. Extending the equation to multiple light sources i, and including the timing information, we can write:

$$\ln \mathcal{L} = \sum_i \sum_j \sum_t k_{jt} \ln(E_i\Lambda_{ijt} + \rho_{jt}) - (E_i\Lambda_{ijt} + \rho_{jt}) - \ln(k_{jt}!). \tag{3.7}$$

Note that the likelihood includes all, also unhit DOMs. Obviously the likelihood depends on the source hypothesis. For amplitude only fits, the timing is ignored, and there is only one time bin spanning the entire event. The hypothesis is $\vec{h} = (x_s, y_s, z_s, t_s, \theta_s, \phi_s, E_s)$ where at time t_s the source with energy E_s at cartesian coordinates x_s, y_s, z_s moving in the direction θ_s, ϕ_s emits photons.

For all event hypotheses, a common framework can be used, the `millipede` framework. It consists of fitting algorithms designed for tracks, single cascades, and double cascades. The `millipede` algorithms can treat noise hits on DOMs. For the single cascade fitting algorithm `monopod` [119], the hypothesis is a single light source. Thus, the minimized parameters are the cascade deposited energy, the incoming neutrino's direction (azimuth, zenith), the cascade vertex position (x, y, z) as well as the vertex time, $\vec{h} = (x, y, z, t, \theta, \phi, E)$. Note that the reconstructed vertex always refers to the shower maximum, i.e. the maximum of the longitudinal energy loss profile. The shower maximum is displaced from the interaction vertex by a few meters in the energy range considered here.

For the track fitting algorithm `mumillipede`, the hypothesis is that of light sources in form of electromagnetic showers being deposited in equally spaced bins along an infinite track. Thus the hypothesis is $\vec{h} = (x_i, y_i, z_i, t_i, \theta, \phi, E_i)$, with the condition $|\vec{x}_{i+1} - \vec{x}_i| = \Delta_{\text{segment}}$, $t_{i+1} - t_i = c\Delta_{\text{segment}}$ imposed, where $\vec{x}_i = (x_i, y_i, z_i)$, corresponding to the equal binning and the particle traveling with the speed of light c and thus faster than the speed of light in ice v_{ice}.

Finally, the double cascade fitting algorithm `taupede` [36] maximizes the likelihood for two causally connected energy depositions. It can be seen as an enhanced single cascade

algorithm, extended by another cascade. This means that there are two additional fitting parameters – the second cascade's energy and the distance between the cascades. The second cascade's direction is the same as the first one's and the second cascade's vertex is uniquely determined by the first cascade's vertex, direction and the double cascade length L_{dc}, i.e. the distance between the two cascades. The parameters to be varied for the double cascade hypothesis are $\vec{h} = (x_1, y_1, z_1, t_1, \theta, \phi, E_1, L_{\mathrm{dc}}, E_2)$, as the tau is traveling in the same direction as the incoming neutrino due to Lorentz boosting. The double cascade separation L_{dc} is a proxy for the tau decay length L_τ. In practice, the light yield and timing at each DOM is compared to the light yield and timing expected from the two energy depositions making up the hypothesis. For the time of the second cascade a similar condition is valid as for track segments: $|\vec{x}_2 - \vec{x}_1| = L_{\mathrm{dc}}$, $t_2 - t_1 = cL_{\mathrm{dc}}$, with $\vec{x}_i = (x_i, y_i, z_i)$. This corresponds to the tau traveling through the ice with the speed of light c.

`Millipede` offers different settings that govern how photons are binned in time, how big the steps in the varied parameters are, or what modules are excluded. The settings also include which ice model is used for reconstruction, which governs how many photons are expected to reach a DOM from a certain distance and direction, and how large the expected time delay due to scattering is wrt. unscattered photons. There are several reasons for the exclusion of a DOM, i.e. its pulses are ignored even if present. DOMs known to take bad or no data, and DOMs with a calibration error are excluded, as there either is no data, or the data cannot be trusted. Then it is possible that an interaction happens very close to a single DOM, and this DOM's signal then accounts for a major part of the total photoelectrons registered from the event. This is problematic, because the more light a DOM records, the more it will contribute to the likelihood in Equation 3.7. However, systematic errors are not included in the likelihood, and the statistical errors become very small for large photon counts. Thus the brightest DOMs may completely dominate the likelihood and bias the reconstruction. For this reason very high energy events with energies in the PeV range cannot be fit well when all modules are considered [124]. The **bright DOM** term was introduced for DOMs that have collected 10 times more light than the average DOM for an event, and data on those DOMs are typically ignored in high energy analyses. Further, a DOM's PMT can also become saturated, leading to the same effects as bright DOMs, with an additional problem of not knowing to a very high precision what number of PE a saturated waveform corresponds to for each individual DOM. Data on saturated DOMs are ignored. Finally, data on DeepCore DOMs are ignored, as they have a higher quantum efficiency than the bulk

IceCube DOMs, and can also bias the reconstruction.
All algorithms in the `millipede` framework can be run iteratively, with the fit with the highest likelihood chosen as the final output. The performance of `millipede` algorithms depends on the quality of the **seed**, the initial hypothesis $\vec{h}_0$ that is varied until a good match between expected and measured light yield is obtained. This has proven a complication especially for `taupede`, where the seed is a hybrid of a cascade and a track. A good such seed is difficult to find, and the reconstruction is seed-dependent. This dependence can be lowered by varying the seed or increasing the number of iterations, both of which come with an increase of computation costs. For `taupede`, an approach was chosen which includes reconstructions on multiple seeds [106]. This will be discussed in the next chapter in more detail.

Seedless, **first-guess algorithms**, are designed to provide a fast first estimate of an event's parameters. They are applied to all real and simulated data and used both for filtering and to provide a seed for more complex algorithms. First-guess algorithms typically only use a small amount of an event's full information, e.g. only the time of the first pulses, the total number of PE registered per DOM, etc. and also assume light propagation through a very simplified, homogeneous ice. A simple first-guess track fitting algorithm is `LineFit` [125], where an initial track is produced on the basis of the times of the first photon arrival at each hit DOM. Assuming a moving light source, and thus a light front traveling with velocity $\vec{v}$ along a straight line through the detector, all hit DOMs can be connected by a line:

$$\vec{r}_i \approx \vec{r} + \vec{v} \cdot t_i \tag{3.8}$$

The fit result is given by the set of $\{\hat{\vec{r}}, \hat{\vec{v}}\}$ that minimizes

$$\chi^2 = \sum_i (\vec{r}_i - (\vec{r} + \vec{v} \cdot t_i))^2 \tag{3.9}$$

where the sum runs over all hit DOMs i with positions $\vec{r}_i$ and photon arrival times t_i. The LineFit result is used as a seed in the `SPEFit` algorithm. The `SPEFit` maximizes the single-photoelectron likelihood, which is given by the arrival times of the first photon registered at each DOM,

$$\ln \mathcal{L} = \sum_j \ln \left(\frac{\mathrm{d}P(\Delta t_{ji} \mid \vec{h})}{\mathrm{d}t} \right), \tag{3.10}$$

where the hypothesis $\vec{h} = (x_i, y_i, z_i, t_i, \theta, \phi)$ is light emitted at (x_i, y_i, z_i, t_i) from a particle moving at the speed of light, and $\mathrm{d}P(\Delta t, \vec{h})/\mathrm{d}t$ is the probability distribution of the photon arrival time [126]. Δt describes the time delay between the geometric time, i.e. the time after which the photon should have arrived at the DOM j from the hypothesized light source i in the absence of scattering, and the observed photon arrival time.
For cascades, a simple maximum-likelihood algorithm is `CascadeLLhVertexFit` [127], which assumes light traveling from a single stationary light source and yields the interaction vertex, with position $\hat{\vec{r}}$ and interaction time $\hat{t}$. `CascadeLLhVertexFit` is used as a seed for `monopod`.
Typically, the low-level algorithms need pulses cleaned from noise, while the `millipede` family of algorithms has a noise term built in and works on raw, uncleaned pulses.
As is explained later on in Section 4.2, HESE is a simple but very effective event selection not relying on the outcome of reconstruction algorithms. All events passing the HESE event selection are first reconstructed with the simple first-guess reconstruction algorithms `LineFit` and `CascadeLLhVertexFit`, as well as `SPEFit` seeded with `LineFit`. The outputs of `CascadeLLhVertexFit`, and `SPEFit` are then used as seeds for the maximum likelihood algorithm `monopod`, the output of which is in turn used to seed `taupede`. The `SPEFit` is the highest-level track fitting algorithm used. The track algorithm from `millipede`, `mumillipede`, is used only to evaluate the energy losses along a given track, without a full minimization.

Chapter 4

High-Energy Neutrinos in IceCube

As described in Section 2.4, a neutrino can interact via NC or CC interaction. When interacting via CC, the interacting neutrino transfers some amount of its energy to the hadron, and some to the charged lepton it creates. The hadron will then produce secondary particles, getting rid of the excess energy, and creating a hadronic cascade in the ice. The charged lepton will also lose energy as it propagates through the ice. Based on the charged leptons' different lifetimes and radiation energy loss profiles, they create potentially distinguishable topologies that can be used to identify the interacting neutrinos' flavors. The main topologies that are created by high-energy neutrinos interacting in IceCube is discussed in Section 4.1. The High-Energy Starting Event selection is introduced in Section 4.2 with a description of the updates developed for and applied to the HESE sample with a livetime of approximately 7.5 years. An overview of all analyses performed on this sample is presented in Section 4.3.

4.1 High-Energy Event Topologies

For high-energy neutrinos interacting in IceCube, there are three main event topologies shown in Figure 4.1 to be distinguished and matched to the three neutrino flavors for a flavor composition measurement.

A **single cascade** consists of one energy deposition and is produced by ν_e-CC and all flavor neutral current (NC) as well as the majority of $\bar{\nu}_e$-GR interactions. In a NC interaction, a hadronic cascade is created. The neutrino only transfers a fraction of its energy, and subsequently leaves the detector. In a ν_e-CC interaction, an electron or positron is created in addition to a hadronic cascade. As described in Section 2.4, the electron promptly radiates off its excess energy, leading to an electromagnetic cascade

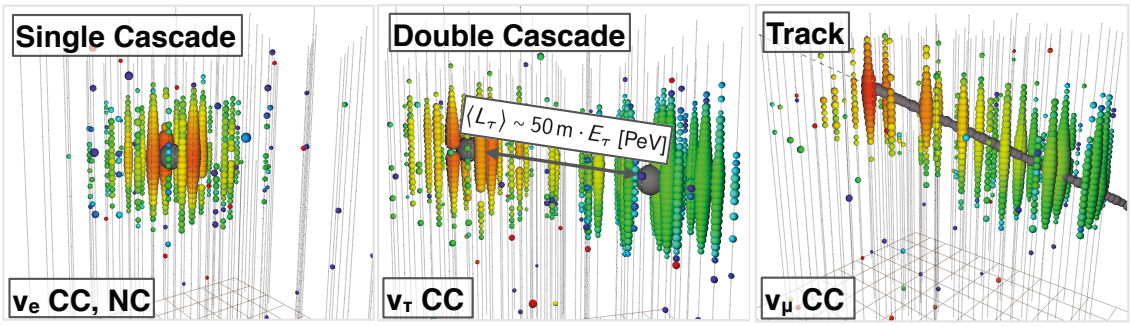

Figure 4.1: Simulated high-energy event topologies in IceCube. A single cascade, made by ν_e-CC and all NC interactions (*left*). A double cascade, created by a ν_τ-CC interaction with the subsequent tau decay producing a hadronic or electromagnetic cascade (*center*). A track is the result of a ν_μ-CC interaction, a ν_τ-CC interaction with the subsequent decay of the tau into a muon, or one or several atmospheric muons (*right*).

overlaid on top of the hadronic one. The cascades' extension is a few meters at most, thus in IceCube, recalling the mean spacing of 125 m between DOMs in the horizontal plane, the light source is effectively point-like[1]. Although the secondary particles are strongly boosted in the direction of their parent particle, the angular information gets washed out quickly due to the photons being scattered before arriving at DOMs. This leads to a poor median angular resolution, which in the HESE sample is $\sim 6°$ for the zenith angle, and $\sim 7°$ for the azimuth angle, excluding systematic uncertainties. In the CC interaction of a ν_e, almost all of the neutrino's primary energy is converted to light: visible energy $\sim$ total deposited energy $\sim$ true neutrino energy. Together with a calorimetric measurement for **contained cascades**, where the interaction vertices are inside the instrumented volume, this leads to a very good median energy resolution, which in the HESE sample is $\sim 8\%$ on the electromagnetic-equivalent deposited energy and $\sim 11\%$ for the true neutrino energy [91] excluding systematic uncertainties. In contrast, the **uncontained cascades**, that are not part of the sample described here, have interaction vertices outside the instrumented volume, and typically a worse resolution in all observables.

Light depositions along a track traversing the detector are called **tracks** and stem from ν_μ-CC interactions and atmospheric muons, as well as ν_τ-CC interactions where the τ lepton decays to a muon. At the energies relevant for this work, the muon's decay

[1] Above ~ 10 PeV, the LPM-effect leads to an elongation of the cascade, making it more egg-shaped. While the elongation might lead to a rise in misclassified single cascades as double cascades, the elongated single cascade would be more than an order of magnitude shorter than a ν_τ-CC interaction at the same energy, and further have a smooth light arrival pattern.

length L_μ surpasses the size of IceCube, and the muon deposits energy mainly in form of stochastic radiative losses on its way through the detector, see Section 2.4. For event selections relying on the number of collected PE, tacks have a higher energy threshold than cascades, as muons only lose a fraction of its energy while traversing the detector. Tracks are further subdivided into **through-going tracks** and **starting tracks**. The former have neutrino interaction vertices outside the detector and are used in the through-going muon analysis. Starting tracks have the neutrino interaction vertex contained within the detector volume, such that the hadronic cascade and part of the muon track are both observed. They are part of the HESE selection described in more detail in the next chapter. Due to the long lever arm, often several 100 m, tracks offer a very good angular resolution down to the sub-degree scale and are therefore used for source searches. The starting tracks in HESE have a median zenith resolution of 1.5°, and a median azimuth resolution of 1.8°. However, the deposited energy is only a lower bound on the estimated neutrino energy, as the muon leaves the detector with a significant fraction of its energy, making spectral measurements more challenging. This is even more difficult for through-going tracks, as the hadronic cascade is not observed, and thus the amount of energy transferred to the nucleus cannot be reconstructed. The tracks in the HESE sample have a median resolution on the deposited electromagnetic-equivalent energy of $\sim 11\%$ [91]. The quoted resolutions do not include systematic uncertainties.
In case of a ν_τ-CC interaction, the tau mostly decays before leaving the detector volume, depositing most of its energy in the decay process. The tau decay length scales linearly with energy, and is on average $\langle L_\tau(E_\tau)\rangle \approx 50$ m E_τ PeV^{-1}. In the energy range accessible to HESE, the tau carries on average 70% of the parent neutrino's energy [128] and transfers on average 30 % of its energy to the tau neutrino it creates upon decay [129]. There are however large event-by-event fluctuations in the amounts of transferred energy and in the decay lengths. If the tau decay produces a muon, the event's topology will be a (starting) track. Naively, one might expect the track to be dimmer than expected from a ν_μ-CC interaction producing a hadronic cascade with the observed light output, because the tau transfers some of its energy to the final state neutrinos. At high enough energies, one might also expect to see a dim track created by the tau, followed by a much brighter track created by the muon, the "sudden" brightening of the track could then serve as indication of a ν_τ-CC interaction. In practice, however, the stochasticity of muon energy losses makes a search for this signature more than challenging, which further suffers from the low branching fraction of the tau decay to a muon of 17.4%. Therefore, we focus on all other decays of the tau. If the tau decays into hadrons, the hadrons will again create a hadronic cascade. If the tau decays producing an electron,

the electron will create an electromagnetic cascade. In both cases we thus have two cascades separated by a distance that scales linearly with the deposited energy. Learned and Pakvasa realized in 1995 [130] that in a large volume neutrino detector a multi-PeV ν_τ-CC interaction can create two cascades separated by several hundred meters and with vertices in the vicinity of different strings of optical sensors, and called this topology unique to ν_τ-CC interactions a **double bang**. A ν_τ interaction creating such a long-lived tau would be identifiable by eye, in line with the by-eye classification of events into tracks and cascades employed in the first iterations of the HESE selection. For the remainder of this thesis, we will refer to all ν_τ-CC interactions where the tau creates a cascade upon decay, as double bangs, even those with very short tau decay lengths.
For a double bang to be identifiable bye eye in IceCube, the tau would need to travel for more than a hundred meters through the detector before decaying, requiring an energy in the PeV range. By May 2018, when the analysis presented here was unblinded, IceCube had seen a total of 6 neutrino events with a reconstructed deposited energy above 1 PeV, including three cascades in HESE. Statistically, the expectation would be that one of the HESE PeV cascades is from a ν_τ-CC interaction. But even with an expected decay length of ~ 50 m, the double bang would not be identifiable by eye due to IceCube's large DOM spacing. To lower the L_τ threshold on ν_τ-CC interactions, two methods have been developed.
If the ν_τ interacts close to a DOM, the interaction and tau decay cascade can be individually resolved by the nearest DOM(s) by analyzing the shape of the waveform. A single energy deposition creates a single "pulse"; the waveform shows a rise, peak, and fall. Two energy depositions close together in space and time can create a **double pulse**; the waveform shows a rise, peak and fall from the first energy deposition whose light reaches the module, and another rise, peak and fall from the second energy deposition. This method focuses on the few brightest DOMs and is most sensitive to ν_τ interactions with $L_\tau \sim 20 - 30$ m.
The other method is to fully reconstruct the created topology of a **double cascade**, using the entire information collected by all DOMs. Double cascades are two energy depositions connected in space and time by a (dim) track. They are produced by double-bang-like ν_τ-CC interactions. As the tau's radiation length is much larger than the muon's, the tau typically only deposits a small, negligible[2] fraction of its energy along its track before it decays. Thus, the description of just two energy depositions is sufficient. Double cascades have a smaller light yield than ν_e-CC interactions: there is at least

[2]For very long lived taus, the track may be visible. For tau track lengths below $\mathcal{O}(100$ m), the tau's energy losses will be outshined by light from the interaction and decay cascades.

one neutrino in the final state carrying away energy, furthermore in most cases the tau produces a hadronic cascade with a smaller light yield than an electromagnetic one, see Section 3.4.1. The higher the neutrino energy, the longer the tau decay length becomes on average, leading to an improved angular resolution. The energy resolution still stays reasonably good (with the already mentioned caveats of invisible energy contributions) as long as the event is contained within the detector volume. Not accounting for systematic uncertainties, the median resolution of the deposited electromagnetic-equivalent energy for double cascades in the HESE sample is $\sim 8\%$, like for the single cascades[3], while the median zenith and azimuth resolutions are 4.4° and 4.7° respectively, slightly better than for single cascades. With IceCube spanning ∼1km in each dimension, the rate of double cascades with both vertices contained in the detector volume drops off above $\sim 300 - 400$ m tau decay length, as the containment becomes more and more constraining to the event's geometry. The short tau decay length of $\langle L_\tau \rangle \sim 50 \text{ m} \cdot E_\tau$ / PeV, where E_τ is the tau lepton's energy, makes the distinction between single and double cascades very challenging in IceCube, given the mean horizontal distance between sensors of 125 m. In practice, the double cascade topology opens up at deposited energies $\mathcal{O}(100 \text{ TeV})$, breaking the degeneracy between ν_e and ν_τ flavors present at lower energies. Using the dedicated reconstruction algorithm `taupede` described in Section 3.4.3, ν_τ events can be reconstructed down to tau decay lengths of just a few meters, making this technique very powerful. The median length resolution is ~ 2 m, owing to IceCube's excellent timing resolution.

Further topologies are possible, but play no role in the analysis presented here. **Stopping tracks** could be created by low-energy muons, or ν_τ-CC interactions where the neutrino interaction vertex is beyond the detector volume. In the latter case, the topology is also called **"lollipop"** due to the cascade created by the decaying tau. Finally, $\bar{\nu}_e$-GR interactions can create a bare track without a hadronic cascade from the decay of the $W^\pm$-boson into a muon. This would also be a starting track, but a **starting bare track**, potentially distinguishable from the starting tracks due to the missing initial cascade.

4.2 The High-Energy Starting Event Selection

The HESE selection is a simple, all-flavor and all-sky selection for high-energy astrophysical and atmospheric neutrinos. It is the event selection with which astrophysical

[3] Almost all double cascades expected in the HESE sample are at double cascade lengths < 100 m, with both cascades contained. This makes them very similar to single cascades.

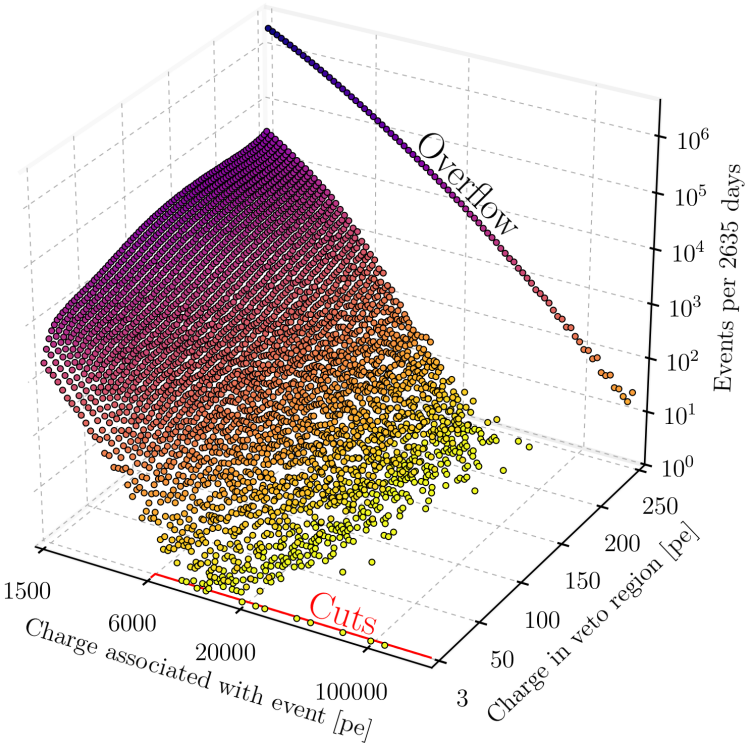

Figure 4.2: Total deposited charge versus charge deposited in the veto region. The color is indicative of the number of events, with darker colors reflecting more events. On the x-axis is the total charge of the events, on the y-axis the charge deposited in the veto region are shown. The "cut" selecting events with less than 3PE deposited in the veto layer is indicated, clearly showing the emergence of a neutrino sample with high purity. Figure taken from [91].

neutrinos were discovered and their flux first measured [23, 88]. Subsequently more data were added, leading to updated results [131, 132]. The selection relies on only two criteria: the registered number of PE called **total charge** Q_{tot}, and a **veto** passing requirement. These selection criteria are chosen to reduce the contribution from atmospheric muons in the sample, while keeping as many high-energy neutrinos as possible. The effect of both criteria is shown in Figure 4.2, where the emergence of a neutrino sample above the muon background is easily visible.

The **veto** region is defined by the outer layer of optical modules, as well as modules in the top 90 m, the bottom layer and in the 60 m wide dust layer in the center. The **fiducial volume** is the volume inwards of the veto, from which events are accepted. The HESE veto is shown in Figure 4.3 with veto DOMs marked in red and DOMs that are part of HESE's fiducial volume marked in blue. The **event start time** is the time

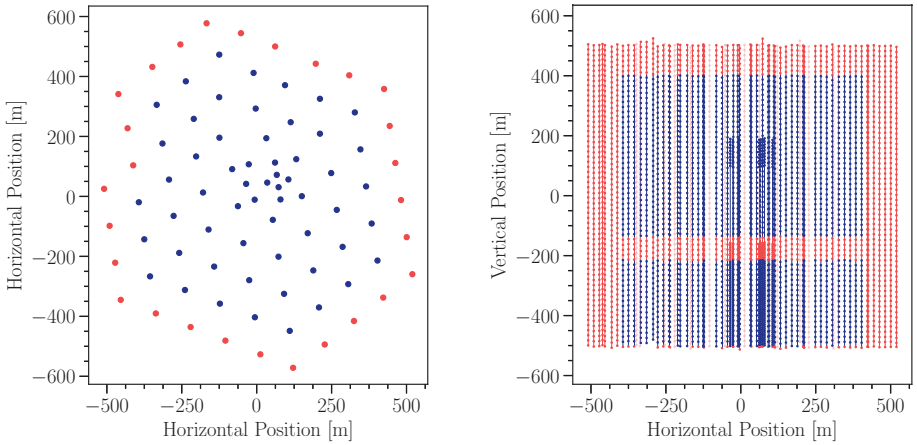

Figure 4.3: Top view (*left*) and side view (*right*) of the veto DOMs and the DOMs in HESE's fiducial volume. The DOMs marked in red are part of the veto, the DOMs marked in blue form the fiducial volume. Figure modified from [91].

after which a total charge of 250 PE is collected from HLC hits throughout the entire detector, excluding DeepCore DOMs. The **event vertex** is not reconstructed at this early stage, but approximated by the charge-weighted positions of the hits contributing to the first 250 PE. An event passes the veto, if this approximate interaction vertex is inside the fiducial volume, and at most 3 PE of the first 250 PE, that are consistent with having been emitted from the approximate interaction vertex, are deposited on the veto DOMs. This selection criterion is specifically designed to discriminate between incoming high-energy muons and starting neutrinos. The veto also reduces the amount of atmospheric neutrinos in the sample, as a fraction of atmospheric neutrinos is accompanied by muons resulting from the same primary CR interaction. This is called the **atmospheric neutrino self-veto**, which is described in Section 4.2.2. Figure 4.4 visualizes how the veto works for three classes of downgoing events: atmospheric muons, atmospheric neutrinos accompanied by atmospheric muons, and solitary atmospheric or astrophysical neutrinos.

To ensure that only high-energy events get selected, a **total charge** of $Q_{\text{tot}} \geq 6000$ PE is required. At the required amount of deposited light, the probability for an incoming muon to deposit light as it enters the detector is high. As the number of PE reaching a DOM is highly dependent on the ice optical properties, it is difficult to translate Q_{tot} into a deposited energy. The energy threshold is higher for events in the upper part of the detector where the ice is less clear, and lower in the deep clear ice. Further, as muons radiate energy over a long path, and typically leave the detector still carrying the majority of their energy, the neutrino energy threshold is also higher for starting ν_μ

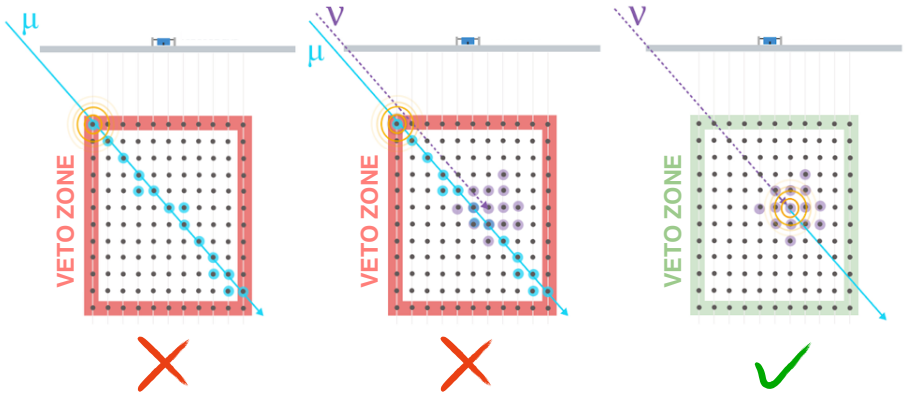

Figure 4.4: The effect of the veto on different high-energy event types: atmospheric muons deposit light in the veto region and do not pass the HESE event selection (*left*), atmospheric neutrinos that are accompanied by muons are also vetoed due to the muons' light deposition in the veto region (*center*), astrophysical and solitary atmospheric neutrinos that interact in the fiducial volume pass the veto region without depositing light (*right*). Figure modified from [95].

interactions than for ν_e interactions. ν_τ interactions also have a higher energy threshold than ν_e interactions, due to the dimmer hadronic interaction cascade and the final state neutrino. The effective area is the area that an idealized, 100% efficient detector with a 100% efficient event selection would have. The effective area for HESE to each of the neutrino flavors is shown in Figure 4.5.

The strengths of the HESE selection clearly lie in its simplicity. The selection does not rely on complex reconstructions, and is thus very robust to e.g. improving filtering or reconstruction algorithms. It is also a high-energy all-flavor, multi-topology, and all-sky event selection. Those properties are very important for this analysis. Of particular importance is that HESE is **all-flavor**: If one wants to measure the flavor composition, one needs a sample that has contributions from all flavors. Further, a handle for each flavor is needed, here this is the event topology. Only a **multi-topology** sample can provide enough information to break degeneracies caused by multiple flavors being able to create the same topology. Finally, the information of the arrival direction of events accessible with an **all-sky** selection enables studies of the astrophysical content of a sample, as atmospheric and astrophysical neutrinos have different zenith distributions.

But HESE also has weaknesses. The high-energy threshold makes an assessment of the background contributions in the analysis region difficult to estimate and leads to high uncertainties on the inferred astrophysical spectral parameters. Further, the employ-

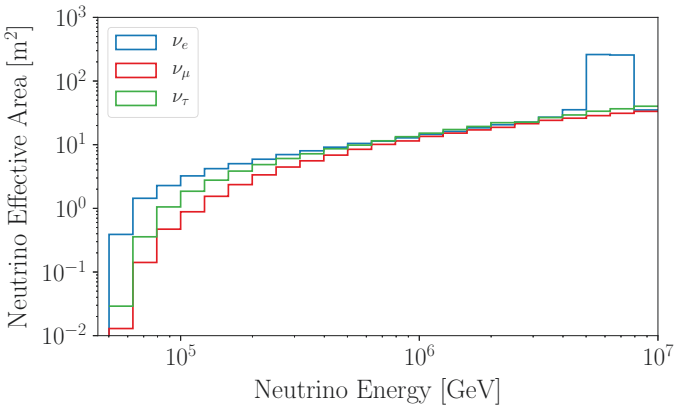

Figure 4.5: Effective area per neutrino flavor for HESE as a function of neutrino energy. The peak at ~ 6 PeV shows the contribution from the Glashow Resonance, highly increasing the $\bar{\nu}_e\, e^-$ cross section.

ment of the veto and the zenith distribution of atmospheric neutrinos, needed for the analysis of astrophysical neutrinos, relies on detailed knowledge and modeling of the atmospheric contributions. An extension of HESE to lower energies, called medium-energy starting events (MESE) [25], has been developed, but has its own difficulties: The muons' probability to deposit light on the outermost DOMs depends on their energy, which led to the development of a dynamic veto region, resulting in a lower fiducial volume at lower energies. However, the weaknesses described above affect the analyses and the interpretation of the data, not the selection. It is therefore possible to apply new methods to the existing selection without changing the content of HESE. There is one exception: as the HESE selection relies on Q_{tot}, i.e. the number of collected PE, a new detector calibration can have a major impact on the selection. This happened for the HESE selection with a livetime of approximately 7.5 years, which will be described next.

4.2.1 What's New in HESE-7

Reprocessing all data to pass-2:
Following the realization that the single photon electron (SPE) peak was misplaced by 5%, a new detector calibration was performed. The SPE peak governs how much current in the PMT corresponds to one PE. This in turn affects the calculation of Q_{tot}, the total number of PE, for all data events. After the SPE correction, all previously recorded data used in HESE were reprocessed from raw waveforms. The new calibration and the data processed with it are referred to as *pass-2*. As the deposited energy is reconstructed

from the registered PE at the DOMs, events get shifted down in energy in *pass-2*, by $\sim 5\%$ averaged over the entire detector. Since HESE analyses have a threshold on the reconstructed total deposited electromagnetic-equivalent energy, $E_{\text{tot}} \geq 60$ TeV, some events will still be part of the sample, but removed from the analyses. Other events will be removed from the sample, as they will fail the $Q_{\text{tot}} \geq 6000$ PE requirement following the reprocessing. Reprocessing was performed applying the best knowledge of the ice optical properties, and the latest analysis tools to all events in a consistent way. Now, all events from all data taking years are processed in exactly the same manner, facilitating year-to-year comparisons.

New self-veto calculation:
The measurement of the astrophysical component is very sensitive to the precise knowledge of the atmospheric contributions in the sample. The atmospheric contributions in turn depend on the effect of the self-veto on atmospheric neutrinos. A new calculation was performed, described in more detail in Section 4.2.2.

SAY likelihood:
Previously, the likelihood assumed infinite MC statistics, which due to limited computational power are never achievable. Especially in a binned analysis, certain analysis bins will suffer from low statistics of the underlying MC simulation, which can lead to biases in the estimation of the fitting parameters. A new, binned likelihood was developed [133], from hereon called SAY likelihood, which takes limited MC statistics into account. Each IceCube simulated event also has a weight w_i which depends on the spectral parameters. Using the effective number of MC events $n_{\text{eff}} = (\sum_i w_i)^2 / \sum_i w_i^2$ and the effective weight $w_{\text{eff}} = \sum_i w_i^2 / \sum_i w_i$, the new likelihood also performs well for weighted MC. It further allows for event-by-event reweighting when updating the hypothesis during the fitting process.

GOLEM fitting framework:
The fitting framework `GOLEM` [134] was designed to be applicable to all analyses performed on the HESE-7 sample. It allows for the study of systematic uncertainties, and a fitting of all model parameters. It is based on an initial incomplete development for a fitting framework able to handle multiple high-energy neutrino samples in a global analysis of the astrophysical spectrum [135]. Its design allows for the easy addition of another sample or another analysis. `GOLEM` uses the SAY likelihood.

Improved treatment of systematic uncertainties:
As already mentioned, the new fitting framework and calculation of the atmospheric passing fractions allow for a variation of the systematics related parameters. The systematic uncertainties that are included in HESE-7 and can be fitted from the data, are:

- Uncertainties in the detector response: the angular response of the DOM, $\epsilon_{\text{head-on}}$, which depends on the optical properties of the hole ice, the absolute DOM efficiency ϵ_{DOM}, and the total anisotropy scale a_s. The latter has an impact on the double cascade length L_τ and therefore on the flavor composition measurement.
- Uncertainties in the cosmic-ray flux: the deviation of the cosmic-ray spectral index $\Delta\gamma_{\text{CR}}$ from the model prediction, the muon normalization scale Φ_μ.
- Atmospheric neutrino flux parameter uncertainties: the conventional and prompt atmospheric normalization scales, Φ_{conv} and Φ_{prompt}, respectively, the Kaon-to-pion ratio $R_{K\pi}$, and the neutrino-to-antineutrino ratio $2\nu/(\nu+\bar{\nu})_{\text{atm}}$.

The systematic uncertainty with the largest impact on the event topology classification is the ice anisotropy, it is discussed in more detail in Section 5.4.

Algorithmic topology classification:
Previous iterations of HESE employed a by-eye classification of the event topology and distinguished between (single) cascades and tracks. This binary topology classification does not allow for a full flavor composition analysis. Further, a 10% misclassification was assumed, which has never been corroborated by a MC study. In HESE-7, we include an algorithmic ternary topology classification. It is entirely MC based, and thus allows for both a quantification of the misclassification fraction, as well as a flavor composition measurement. While this topology classification has already been applied to the 6-year HESE sample [37], it is now an inherent part of the sample itself. The classification relies on the iterative reconstruction of the full event topologies. Not only does the classification supersede the by-eye classification, also the reconstruction of the observables are now provided by its underlying iterative reconstruction. The topology classification is fully integrated into `GOLEM`, including the ice anisotropy systematic treatment. It is discussed in detail in the next chapter.

4.2.2 The Atmospheric Neutrino Self-Veto

Recall the description of atmospheric neutrinos in section 2.6: the only particles born in CR-induced showers that can reach sub-surface detectors are neutrinos and muons,

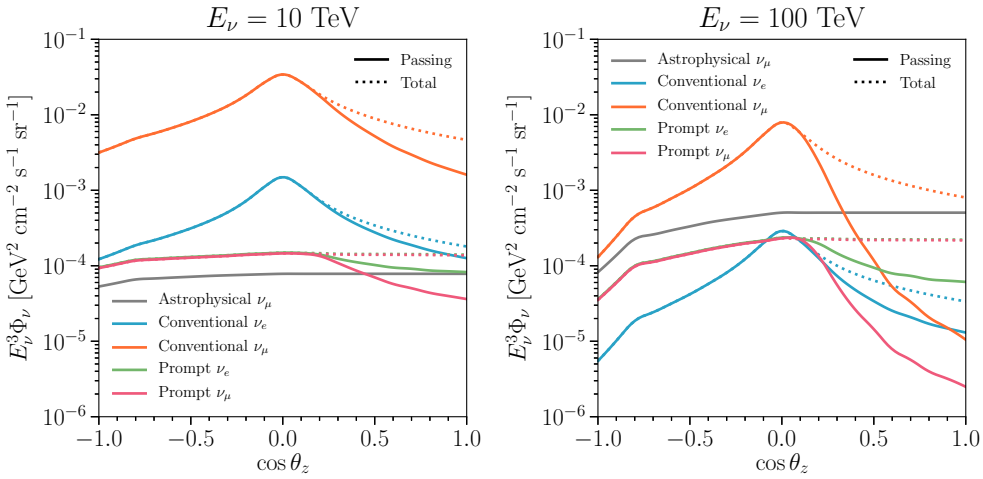

Figure 4.6: Atmospheric neutrino fluxes and the effect of the self-veto, for two neutrino energies, $E_\nu = 10$ TeV (*left*) and $E_\nu = 100$ TeV (*right*), shown as a function of the cosine of the zenith angle. The fraction of the flux not vetoed ("passing") fluxes (solid), and the total fluxes entering the detector (dashed) are shown for several components of the total flux. Figure taken from [136].

and high-energy CR-showers contain a multitude of neutrinos and muons. Thus, every atmospheric neutrino that interacts in IceCube is accompanied by other neutrinos, that typically will not interact at the same time, as well as muons. Depending on the primary cosmic ray's energy and zenith angle, some muons might reach IceCube in coincidence with the interacting neutrino and deposit light on the outer DOMs. Thus, the probability of an atmospheric neutrino being rejected by an accompanying muon triggering the veto depends on its energy and on its zenith angle. No muons reach IceCube through the Earth; the self-veto only works for downgoing neutrinos.

In previous iterations, a step-function was assumed for the probability p_{light} for a muon to trigger the veto when reaching the detector: $p_{\text{light}} = 0$ for $E_\mu < 1$ TeV and $p_{\text{light}} = 1$ for $E_\mu \geq 1$ TeV [137].

Now, a new calculation has been performed [136], in which p_{light} is a smooth function of E_μ reflecting a realistic detector response. The new calculation uses `MCEq`[138] to calculate atmospheric lepton fluxes. `MCEq` provides numerical solutions to the coupled cascade equations describing the development of cosmic-ray induced showers in matrix form, making calculations of atmospheric lepton fluxes precise and computationally efficient. It supports a variety of models of cosmic-ray spectra and hadronic interactions [46].

The baseline cosmic ray model is the Hillas-Gaisser three-population model (H3a) [139]

used to weigh the atmospheric neutrino contributions. The baseline hadronic interaction model is SIBYLL2.3c [140], and the baseline atmospheric density model is MSIS-90E [141, 142].
Other available models have also been tested with the conclusion that the fraction of muon neutrinos passing through the veto depends only very slightly on the underlying models. The model variations have a non-negligible effect on the electron neutrino passing fractions, however, the electron neutrino fluxes are an order of magnitude below the muon neutrino fluxes. The new calculation offers a way to include systematic uncertainties from model variations.
The effect of the veto on the atmospheric muon component μ_{atm}, the conventional atmospheric neutrino component $\nu_{\alpha,\text{conv}}$, the prompt atmospheric neutrino components $\nu_{\alpha,\text{prompt}}$ and the astrophysical muon neutrino component $\nu_{\mu,\text{astro}}$ is shown in Figure 4.6. The passing fraction of atmospheric neutrinos is the fraction of atmospheric neutrinos that are not accompanied by a detectable muon from the same cosmic-ray-induced shower. At $E_\nu > 100$ TeV, the astrophysical component dominates over the atmospheric ones. The astrophysical neutrino components $\nu_{e,\text{astro}}$ and $\nu_{\tau,\text{astro}}$ behave in the same way as the astrophysical muon neutrino component $\nu_{\mu,\text{astro}}$ shown in the figure.

4.3 Analyses Performed on HESE-7

In the previous iterations of HESE, the events were used to characterize the **diffuse astrophysical spectrum assuming a single power-law**, as well as the conventional and prompt atmospheric contributions. This is still a key measurement of the 7.5 year HESE sample. In addition, due to the better statistical treatment and the larger statistics, not only a single power-law can be tested, but also a double power-law, or a potential cut-off in the spectrum. The sample and the diffuse fits are presented in [91].
One major achievement of the unified fitting framework is the ability to include other measurements in a self-consistent way. The **neutrino-nucleon cross-section** was measured with the sample using the Earth absorption [143]. Further, **Dark Matter searches** were performed, looking for signatures imprinted on the neutrino spectrum in the Dark Matter annihilation and the Dark Matter scattering channels [144]. Finally, constraints to various physics mechanisms **beyond** tho ones in the **standard model** (BSM) can be set, which could alter the propagation and flavor composition of neutrinos [145].
The **point source search**, an integral part of HESE analyses from the first iteration, was also performed. It has to be noted, however, that the point source search does not

use the iterative reconstruction that is used for all other analyses. Instead, the entire sky is scanned on a grid of directions and the likelihood is calculated for each direction. This allows to draw a contour for the directional uncertainty of each of the events, which is important when assessing the self-clustering of the events or the closeness to the galactic plane. However, those sky scans are computationally very expensive, and thus unfeasible to perform and test on a large MC sample. Further, the sky is scanned on a finer grid for tracks than for cascades; a double cascade sky scan has not been implemented. The point source search was not updated to use the ternary algorithmic topology identifier used in all other analyses on the sample, instead, it uses the old by-eye classification. The source search is presented together with the diffuse results included in [91].
The **flavor composition analysis** presented in the next chapter assumes a single power-law with the same spectral index for all neutrino flavors. It is an extension of the single power-law fit to the diffuse spectrum by two parameters describing the relative flavor contributions to the total normalization. Thus, with the exception of the single power-law diffuse fit, the other analyses performed on the HESE sample mentioned in this section are not relevant to the remainder of this thesis.

Chapter 5

Flavor Composition Analysis

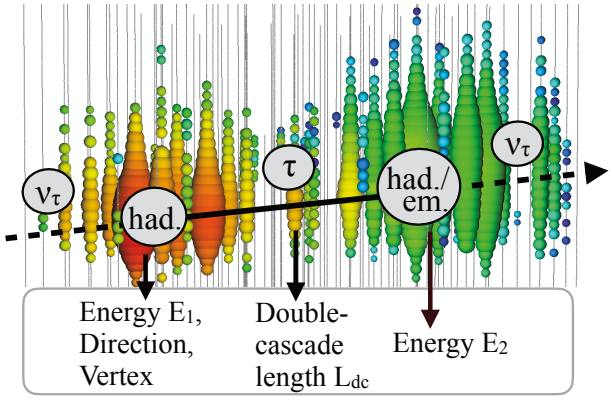

Figure 5.1: Simulated 10 PeV ν_τ-CC interaction in IceCube creating a double cascade. The first cascade stems from the hadronic interaction of the ν_τ, producing a tau lepton, the second cascade stems from the tau decaying, producing either hadrons or an electron which in turn create the hadronic or electromagnetic second cascade. The double cascade observables are shown below: the energy, direction, vertex position and time of the first cascade; the double cascade length from the separation of the two cascades; and the energy of the second cascade.

If a ν_τ interacts via CC interaction, a tau lepton is created, which quickly decays. If the tau is sufficiently long-lived and does not decay into a muon, the ν_τ-CC interaction in IceCube may create a double-cascade signature unique to tau neutrinos. This signature, how it can be used to identify ν_τ interactions, and how the flavor composition of neutrinos can be measured based on their interactions' topologies are the topics of this chapter. Tau neutrinos are the only flavor that can create all three topologies, as NC interactions create single cascades, double-bang-like CC interactions create double

cascades and CC interactions where the tau decays producing a muon create tracks.
At the beginning of this chapter in Section 5.1, the reconstruction of HESE events and selection of double cascades is described, and how the double-cascade properties can be used to classify event topologies, maximizing the ν_τ-content in the selected double cascades. In Section 5.2, the efficiency of the algorithmic topology classification is discussed. The flavor composition measurement using MC templates and the event topologies is described in Section 5.3. At the end of this chapter, major systematic uncertainties on the topology classification are discussed in Section 5.4.

5.1 Reconstruction and Selection of Double Cascades

A double cascade is created when a ν_τ interacts via CC, creates a tau lepton that then decays producing a second decay cascade, i.e. for all tau decay channels which do not produce a muon. We call all such interactions, regardless of the tau decay length L_τ, **true double cascades**, as long as light from both cascades is visible in IceCube. This in principle leads to a geometry-dependent maximum L_τ of a few hundred meters. Events with longer lived taus typically have one of the cascades outside of IceCube, such that we would have a starting or stopping track topology instead. If a double cascade's both vertices are close to a DOM, a double pulse may also be observed.
To reconstruct the full event topology, the maximum likelihood reconstruction algorithm `taupede` is used, which has been described in Section 3.4.3. Recall that `taupede` maximizes the likelihood for the hypothesis of two causally connected energy depositions given the observed light yield, with parameters $\vec{h} = (x_1, y_1, z_1, t_1, \theta, \phi, E_1, L_{\mathrm{dc}}, E_2)$. It is those parameters that are explored for the classification of events as double cascades, single cascades, and tracks. The reconstruction algorithm returns the event hypothesis' parameters with the highest likelihood given the observed data. However, this game can be played for each of the three event hypotheses and matching reconstruction algorithms, yielding three sets of best fit event parameters. To classify the events and assign topologies, one needs to find out which event hypothesis is the best one. For that, the reconstructed observables, as well as the likelihood values of the fits can be used.
The analysis performed here has two objectives: identify double cascades created by tau neutrinos, and measure the flavor composition of astrophysical neutrinos. As the identification of double cascades is very challenging, and the double cascade reconstruction algorithm is prone to failures and dependent on the supplied seed hypothesis, all events are reconstructed with all topology hypotheses. In [106], MC events of known topologies were scrutinized for their typical observable distributions to find a set of reconstructed

properties that maximizes the tau neutrino content in the events classified as double cascades, while maintaining a high purity in the track and single cascade topologies as well. The default hypothesis for the analysis is a double cascade. Following reconstruction, each event's observables are examined for indications of a bad double cascade fit according to the procedure developed in [106] and shown in Figure 5.2, and the final topology is assigned.

All events are reconstructed using L2 cascade and track reconstruction algorithms, and regardless of their outcome, using L3 cascade and track algorithms. Then, an amplitude-only `monopod` fit is performed, seeded with a combined seed created by using the cascade fits for the vertex, and the track fits for the direction. The result of this fit is used as seed for a full `monopod` fit with four iterations. Based on the `monopod` fit, multiple seeds for `taupede` are constructed as it has been difficult to find a reliable seed for `taupede`. They use the `monopod` fit as seeded particle, and length seeds of 10, 25, 50, and 100 m. Further, all seeds are used in forward, center, and backward mode, where the forward and backward seeds are shifted by the seed length along the seed direction. An amplitude only `taupede` fit is performed for each seed. Out of the 12 fits, the three best ones are chosen as seeds for a full `taupede` fit with four iterations, which take the photon arrival times at the DOMs into account. Scattered photons from a single energy deposition have different photon arrival time patterns than photons from two energy depositions which are separated by L_{dc} in space and L_{dc}/c in time. Finally, a `mumillipede` unfolding is performed along the direction of each of the topology fits (`monopod`, `taupede`, `SPEFit16`), and the best fit based on its likelihood value is chosen for final comparison.

The primary **observables** of the double cascade reconstruction algorithm are the interaction vertex x_1, y_1, z_1 and the decay vertex x_2, y_2, z_2, the electromagnetic-equivalent energy of the first and second cascade, E_1 and E_2 respectively, the distance between the vertices L_{dc}, as well as the reduced log likelihood, `rlogl` $= \log(\mathcal{L})/\mathrm{ndof}$ where ndof is the number of degrees of freedom. These observables are used to construct the final observables used for classification of the HESE events. These final observables are:

- The **total deposited energy**, E_{tot}, is the reconstructed electromagnetic-equivalent deposited energy. It is the sum of all `mumillipede` losses and is required to be ≥ 60 TeV for all events classified. Events with a lower reconstructed total deposited energy are discarded from all fits performed with HESE.

- The **direction** is compared for the fit with the highest `mumillipede` likelihood and the `taupede` fit. If these are different, the opening angle between the two need to be $\leq 30^{\circ}$ for a double-cascade preselection.

- The **soft containment** is a binary cut requiring each of the two vertices to be at most 50 m outside of detector boundary, which is defined by the outermost DOMs. A vertex reconstructed far outside the detector volume indicates a true topology different from a double cascade, either a track where a big cascade is placed such that it maximizes the likelihood for the light seen in the detector, or a single cascade, where there is no light anywhere else in the detector. Soft containment is required for preselection.

- The **energies of the individual cascades**, E_1 and E_2, are both required to be ≥ 1 TeV. As this is a high-energy analysis, any event with a resolvable double cascade should have much higher-energy individual cascades. A cascade with very low energy indicates a true single cascade as underlying topology. The minimum energies of 1 TeV each are required for preselection.

- The **double-cascade length** L_{dc} is is a proxy for the tau decay length L_τ and is required to be $L_{\mathrm{dc}} \geq 10$ m for double cascades. Many ν_τ will produce double cascades of shorter length and be classified as single cascades. True single cascades are reconstructed to low lengths, usually ≤ 2 m, with the number of events falling off exponentially at larger lengths. While it decreases the total number of identifiable tau neutrinos, the minimal required length of 10 m keeps the background levels low and thus increases the ν_τ purity in the double-cascade sub-sample.

- The **energy asymmetry**, $A_E = (E_1 - E_2)/(E_1 + E_2)$ is a measure of the inelasticity of the interaction. In general, the distribution of the variable is almost flat for true double cascades, although the analysis is more sensitive to events with negative A_E. A high positive A_E indicates a true single cascade, where most of the energy is given to the first reconstructed cascade, and only little energy to the second reconstructed cascade. $-0.97 \leq A_E \leq 0.3$ is required for double cascades.

- The **energy confinement**, $E_C = (E_{1,C} + E_{2,C})/E_{\mathrm{tot}})$ is constructed using a `mumillipede` unfolding of the energy deposited along the track. $E_{1,C}$ ($E_{1,C}$) is the sum of `mumillipede` losses within 40 m of the first (second) `taupede` vertex. $E_C \geq 0.99$ is required, as a low energy confinement is indicative of a true track with the muon depositing a significant amount of energy along the track either between the two reconstructed cascades, or beyond.

- For events that do not pass the preselection described above, the **reduced log-likelihood** `rlogl` of the track and single-cascade fits are compared to assign the final topology.

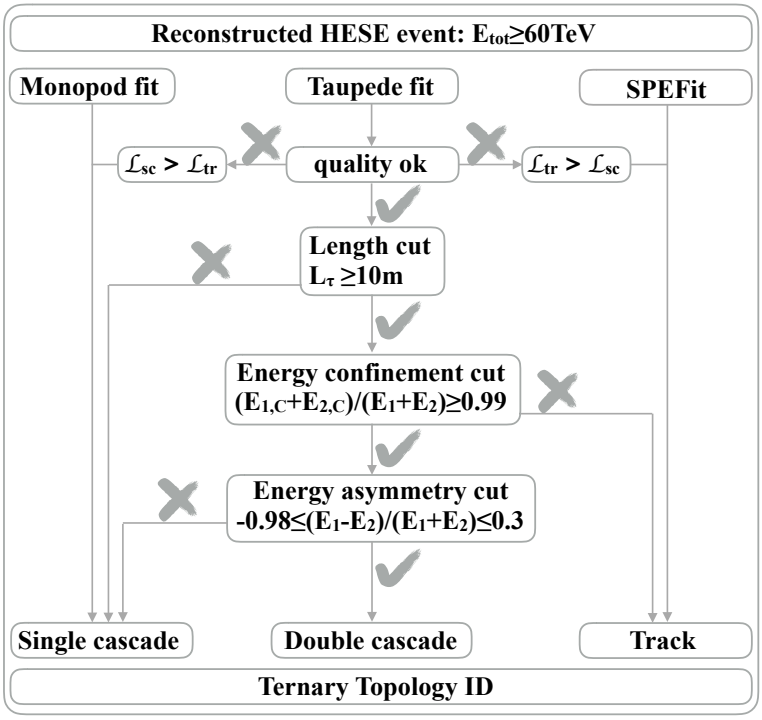

Figure 5.2: Ternary topology classification chain. The first step labelled "preselection" requires the passing of soft containment, meeting the direction and the minimum energies requirements, in addition to the `taupede` fit converging. For events failing the preselection, the likelihoods of the track and single cascade fits are compared, and the topology with the corresponding fit with the higher likelihood is chosen. Note that while there is no upper limit on the length, the requirement that each cascade can be at most 50 m outside of the instrumented volume leads to an effective cutoff above $\sim 300 - 400$ m.

The entire classification procedure is shown in the Figure 5.2. It was developed as a stand-alone analysis that was applied on the HESE sample with 6 years of livetime. The development of the analysis is described in great detail in [106]. Also in the original analysis, all events were processed in a consistent way, assuming the same ice model (SpiceLea at that time). Here, this analysis is re-applied to the events with 7.5 years of livetime passing the HESE selection. The changes and updates wrt. the original analysis are kept to a minimum and necessitated by the reprocessing of data following the *pass-2* detector calibration, the new analysis tools and the enhanced knowledge and treatment of systematics. Further, the analysis is now an integral part of the HESE sample, providing each event with the final reconstructed observables and the topology.

True Topology	Reconstructed Topology: single	double	track	Reconstructed Topology	True Topology: single	double	track
single	69%	24%	12%	single	93%	77%	25%
double	22%	67%	7%	double	2%	16%	2%
track	9%	9%	81%	track	5%	7%	73%
Expected Contributions:							
μ_{atm}	0.0	0.0	0.8	μ_{atm}	0.0	0.0	0.8
ν_{atm}	5.1	0.2	4.4	ν_{atm}	3.4	0.2	6.0
ν_{astro}	55.0	4.0	13.7	ν_{astro}	41.1	17.4	14.3
Total	60.1	4.2	18.9	Total	44.5	17.6	21.1

Table 5.1: True topology vs. reconstructed topology for the baseline astrophysical and atmospheric spectra given in Equation 5.1. The columns sum up to 100%; the entries on the diagonal show the purity of each reconstructed topology, while the off-axis entries show the misidentification fraction (*left*). The entries on the diagonal show the classification efficiency (*right*). The bottom lines give the number of expected events from atmospheric and astrophysical contributions per topology, and the total number of events.

5.2 Topology Classification Efficiency

The ternary topology classification described in Section 5.1 and depicted in Figure 5.2 was developed to maximize the efficiency to identify ν_τ-induced double cascades, as these are both rare and of utmost importance for a measurement of the flavor composition. Therefore, the double cascade is the default event hypothesis. Maximizing the double cascade identification efficiency by construction decreases the sensitivity to single cascade and track identification. However, a comparison between the original analysis [106] and an analysis that developed an algorithmic cascade and track discrimination for starting events with a minimum of 1 TeV deposited energy [146] revealed the same level of misclassified cascades and tracks in both analyses. Thus, the entire analysis chain and topology identification scheme developed for the 6-year HESE tau search and flavor composition measurement is kept here. The astrophysical spectrum, on which all analyses were developed or tested, is a single power-law based on the combined fit to multiple high-energy samples [32],

$$\frac{\mathrm{d}\Phi_{\text{astro}}}{\mathrm{d}E} = 6.9 \cdot \left(\frac{E}{100\ \text{TeV}}\right)^{-2.5} \cdot 10^{-18} \cdot\ \text{GeV}^{-1}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1} \tag{5.1}$$

Flavor	Reconstructed Topology			Recon-structed	Flavor		
	single	double	track	Topology	ν_e	ν_τ	ν_μ
ν_e	57%	20%	10%	single	93%	76%	37%
ν_τ	29%	71%	13%	double	2%	13%	2%
ν_μ	14%	10%	73%	track	5%	11%	61%
μ_{atm}	0.4%	0%	4%				
Expected Contributions:							
μ_{atm}	0.0	0.0	0.8	μ_{atm}	0.0	0.0	0.0
ν_{atm}	5.1	0.2	4.4	ν_{atm}	2.1	0.2	7.5
ν_{astro}	55.0	4.0	13.7	ν_{astro}	34.9	22.7	15.4
Total	60.1	4.2	18.9	Total	37.0	22.5	22.9

Table 5.2: Flavor vs. reconstructed topology for the baseline astrophysical and atmospheric spectra given in Equation 5.1. The columns sum up to 100%; the entries on the diagonal show the flavor identification efficiency (*left*), and the flavor purity in each sub-sample (*right*). See Table 5.1 for details.

with total normalization $\Phi_{\text{astro}} = 6.9 \cdot 10^{-18} \cdot \text{GeV}^{-1}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}$ and spectral index $\gamma_{\text{astro}} = 2.5$, as well as a flavor composition on Earth of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$. The baseline atmospheric spectra are given by the HKKMS [147] calculation for the conventional component and BERSS [148] calculation for the prompt component.

With these assumptions, the expectation is to observe 83.2 events and to classify them into 60.1 single cascades, 4.2 double cascades, and 18.9 tracks. The purity and misclassification ratio is shown in the left part of Table 5.1. From that it follows that 2.8 events are expected to be true ν_τ-induced double cascades, while 1.4 events are expected to be classified as double cascades but stemming from other interactions, including ν_τ-NC interactions. The atmospheric contributions to the double cascade sub-sample come from prompt tau neutrinos and amount to 0.2 events following the $\nu_{\tau,\text{prompt}}$ flux calculations by [46]. The main background for the identification of ν_τ interactions via the double cascade classification algorithm are thus $\nu_{e,\text{astro}}$ and $\nu_{\mu,\text{astro}}$. Recall that there is a mismatch between what constitutes a true double cascade, namely a ν_τ-CC double-bang-like interaction with arbitrarily short tau decay length L_τ, and a reconstructed double cascade, which among others requires a reconstructed double cascade length $L_{\text{dc}} \geq 10$ m. Thus, well-reconstructed true double cascades with $L_\tau \sim L_{\text{dc}} < 10$ m will by construction be classified as single cascades. This explains the low efficiency to classify true double cascades as double cascades shown in the right part of Table 5.1. In total, 17.6 ν_τ events interacting in the double-cascade channel are expected, but only 4.2 of them are identi-

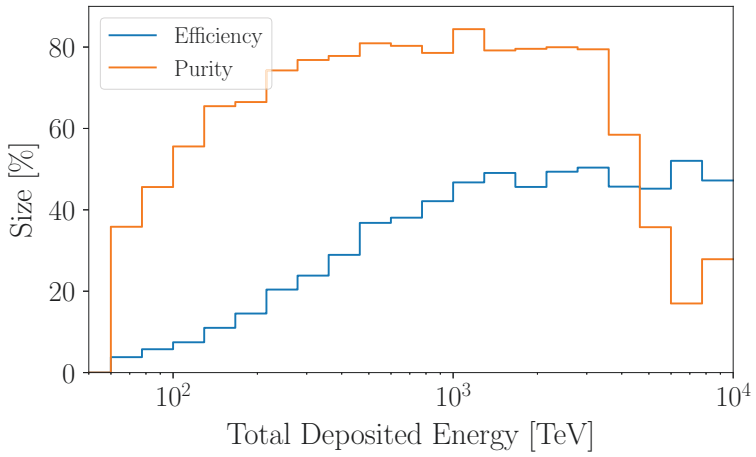

Figure 5.3: Double-cascade classification efficiency and purity as a function of the reconstructed energy.

fiable with the classification procedure shown in the Figure 5.2.
The flavor content of each reconstructed topology sample is shown in Table 5.2. From the left part of the table, it is clear that the reconstructed topology carries information about the interacting neutrino's flavor, but is not a direct proxy for it. The per-flavor normalization on the right part of the table reveals that 76% of ν_τ interactions will be classified as single cascades. Only a fraction of them are NC interactions, the majority being ν_τ-CC interactions producing true double cascades with short tau decay lengths L_τ that fail the $L_{\mathrm{dc}} \geq 10$ m requirement for selected double cascades. The energy asymmetry requirement of $-0.98 \leq A_E \leq 0.3$ leads to a further reduction of true double cascades classified as double cascades. The number of expected ν_e events is largest, owing to the larger effective area of HESE to ν_e interactions shown in Figure 4.5. Only 13% of the expected 22.9 ν_τ events in HESE are expected to be classified as double cascades. The double-cascade classification efficiency and purity are shown in Figure 5.3 as a function of the reconstructed energy. The efficiency is very low at the lower energy bound of the analysis range, and reaches $\sim 20\%$ for $E_{\mathrm{tot}} \sim 200$ TeV. The rapid decrease in the purity around $E_{\mathrm{tot}} \sim 6$ PeV is due to the Glashow resonance for $\bar{\nu}_e$.

5.3 Flavor Composition Measurement

The topology is not simply used as a proxy for the neutrino flavor. Instead, MC templates are used to infer the best flavor composition given the number of events classified into each topology as well as the distributions of their observables. For each flavor and for

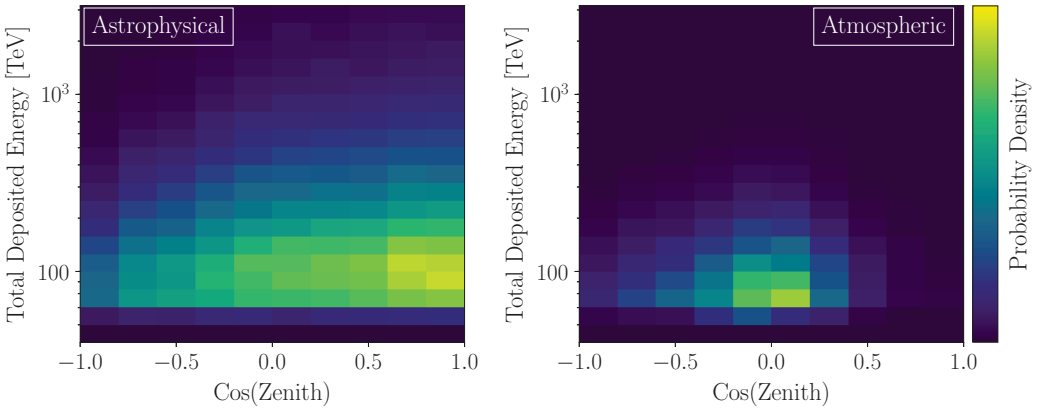

Figure 5.4: Total reconstructed energy versus reconstructed cosine of the zenith angle for events classified as single cascades. The signal, i.e. ν_{astro}-induced, single cascades are mainly downgoing and extend to higher E_{tot} (*left*). The background, i.e. induced by atmospheric events, single cascades mainly come from horizontal directions and cluster at low E_{tot}. The baseline astrophysical and atmospheric spectra described in Section 5.2 are assumed.

each topology, a MC template is created, which is a two-dimensional probability density function (PDF). For single cascades and tracks, the observables total deposited energy E_{tot} and the cosine of the zenith angle $\cos(\theta_z)$ are used, as these observables carry the highest power to discriminate between the astrophysical and atmospheric components and are well established in the previous HESE analyses. The $\cos(\theta_z)$ distribution is expected to be very similar for astrophysical neutrinos of all flavors in the HESE sample. The two-dimensional MC PDF templates used for single cascades and tracks are shown in Figures 5.4 and 5.5, respectively. The left panels show the expected distribution of astrophysical neutrinos, while the right panels show the expected distribution of from the atmospheric contributions, i.e. conventional and prompt neutrinos and for tracks also from atmospheric muons. For double cascades, astrophysical ν_e and ν_μ make up most of the background for the ν_τ search. Thus, the observables E_{tot} and L_{dc} are used, as these two have the highest power to discriminate between the flavors of astrophysical neutrinos. This can be seen from the following argument:
Under the double cascade hypothesis, true single cascades are reconstructed to very small L_{dc} on average, or to large L_{dc} but with most of their energy in the first cascade, with $A_E \to 1$. Requiring the reconstructed length $L_{\text{dc}} \geq L_{\text{dc, threshold}}$ only keeps true single cascades in the tails of the approximately exponentially falling L_{dc} distribution in the double cascade sample. The energy asymmetry selection criterion further reduces the single cascade misclassification probability. Thus, the true single cascades misclassified

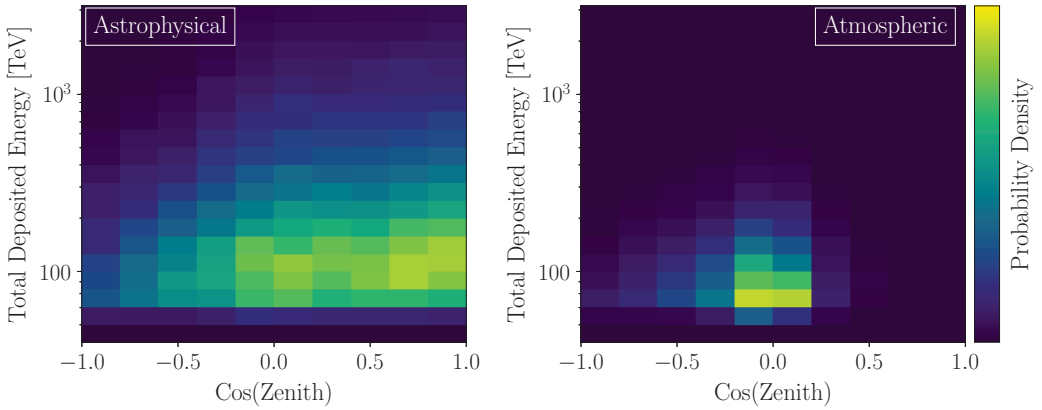

Figure 5.5: Total reconstructed energy versus reconstructed cosine of the zenith angle for events classified as tracks. See the caption of Figure 5.4 for details.

as double cascades cluster at L_{dc} close to the threshold defined by the analysis, which here is 10 m. Their E_{tot} distribution will follow the energy distribution of astrophysical neutrinos. We thus expect misclassified true single cascades to cluster at low L_{dc} and low E_{tot}, the L_{dc} distribution falling off more rapidly.

True tracks have many energy depositions as the muon loses energy in stochastic losses while traversing the detector. Reconstructing an arbitrary number of cascades with a two-cascade hypothesis will pick the regions of the two largest energy depositions as the resulting cascades. Thus, true tracks will be reconstructed to arbitrary L_{dc}. However, there will be energy depositions along the tracks but far from the vertices, leading to the energy confinement $E_C < 0.99$ in most cases. The tracks misclassified as double cascades typically cluster at low E_{tot}. This is partly due to the falling astrophysical spectrum and the higher energy threshold for tracks in HESE. Further, higher-energy muons on average produce more large energy losses along their path, leading to them failing the energy confinement criterion.

True double cascades from ν_τ-CC interactions will lie in a special region on the $L_{\mathrm{dc}} - E_{\mathrm{tot}}$ diagram. The average tau decay length scales with its energy, $\langle L_\tau \rangle \sim 50\ \mathrm{m} \cdot E_\tau/\mathrm{PeV}$, where $\langle E_\tau \rangle \sim 0.7\, E_{\nu_\tau}$, albeit with large event-to-event variations. Thus the distribution of true double cascades is expected to show approximate proportionality between the L_{dc} and the E_{tot} observables. This correlation is mainly used to assess how compatible a double cascade is with a ν_τ versus another flavor interaction. The MC PDFs for double cascades are shown in Figure 5.6. The correlation between the reconstructed total deposited energy E_{tot} and the reconstructed double cascade length L_{dc} is clearly visible for the ν_τ-induced double cascades shown in the left panel. On the right panel,

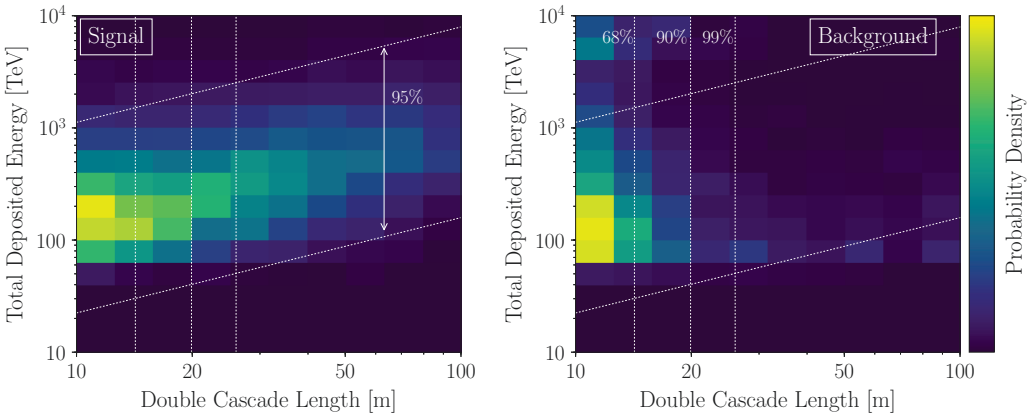

Figure 5.6: Total reconstructed energy versus reconstructed double cascade length for events classified as double cascades. The signal (= ν_τ-induced) double cascades show a correlation between L_{dc} and E_{tot}, with 95 % of events lying in the indicated signal region (*left*). The background (=ν_μ and ν_e-induced) contributions do not show this correlation, and cluster at low L_{dc} and E_{tot}. The vertical lines indicate the reconstructed length below which a given percentage of background from true single cascades is contained (*right*). The baseline astrophysical and atmospheric spectra described in Section 5.2 are assumed.

the background distribution follows the patterns explained above.

Using these MC templates for double cascades and the analogous ones for single cascades and tracks with the observables $\cos(\theta_z)$ and E_{tot}, the flavor composition is fit by scaling the contributions of each flavor. This is done in addition to varying the other parameters of the fit, and re-weighting the MC events accordingly for each variation. The parameters of interest are the relative flavor contributions f_α, $\alpha = e, \mu, \tau$, the astrophysical spectral index γ_{astro}, and the total astrophysical flux normalization Φ_{astro}. The nuisance parameters are the parameters Φ_{conv}, Φ_{prompt}, Φ_μ, $\Delta\gamma_{\mathrm{CR}}$, $R_{K\pi}$, $2\nu/(\nu+\bar{\nu})_{\mathrm{atm}}$, ϵ_{DOM}, $\epsilon_{\text{head-on}}$, and a_s described in Section 4.2.1. The priors and valid ranges for all model parameters are given in Table 5.3. Φ_{conv} is given in units of the predicted flux by the HKKMS calculation [147], Φ_{prompt} is given in units of the predicted flux by the BERSS calculation [148], $R_{K\pi}$ in units of the predicted ratio by the MR calculation [149]. Φ_μ is in units of the HGH4a flux calculated by [150], with a modification of the normalization derived from data. This treatment allows deviations on the normalization of the flux, with the shape being fixed by the calculations. For the detector systematics, variations of ϵ_{DOM} describe a total scaling of all DOM efficiencies relative to the baseline individual DOM efficiency, $\epsilon_{\text{head-on}}$ scales the angular acceptance of the DOMs relative

Parameter	Constraint	Range	Description
Φ_{astro} $[10^{-18} \cdot \text{GeV}^{-1}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}]$	-	$[0, \infty)$	Normalization
γ_{astro}	-	$(-\infty, \infty)$	Spectral index
f_α	$\sum_\alpha f_\alpha = 1$	$[0, 1]$	Relative flavor contributions
Φ_{conv} [HKKMS]	1.0 ± 0.4	$[0, \infty)$	Conventional normalization
Φ_{prompt} [BERSS]	-	$[0, \infty)$	Prompt normalization
$R_{K\pi}$ [MR]	1.0 ± 0.1	$[0, \infty)$	Kaon-to-pion ratio
$2\nu/(\nu + \bar{\nu})_{\text{atm}}$	1.0 ± 0.1	$[0, 2]$	Neutrino-anti-neutrino ratio
$\Delta\gamma_{\text{CR}}$	0.0 ± 0.05	$(-\infty, \infty)$	Cosmic-ray spectral index
Φ_μ $[2.1 \cdot (N_{\text{tagged}}/N_{\text{expected}})\cdot$ HGH4a]	1.0 ± 0.5	$[0, \infty)$	Muon normalization
ϵ_{DOM} [DOM model]	0.99 ± 0.1	$[0.80, 1.25]$	Absolute DOM efficiency
$\epsilon_{\text{head-on}}$ [ice model]	0.0 ± 0.5	$[-3.82, 2.18]$	DOM angular response
a_s [ice model]	1.0 ± 0.2	$[0.0, 2.0]$	Ice anisotropy scale

Table 5.3: Analysis model parameters for the single power-law flavor composition fit. Constraints for analysis parameters are either uniform or Gaussian. The ranges give the physically allowed values of the parameters.

to the baseline hole-ice parameters and a_s scales the strength of the anisotropy relative to the values of the *Spice3.2* ice model of 10% for the major axis and 5% for the minor axes. The systematic uncertainties are described in more detail in the next section.
The likelihood used is the SAY likelihood $\mathcal{L}_{\text{Eff}}$ [133], given by

$$\mathcal{L}(\vec{\theta}, \vec{\xi}) = \mathcal{L}^{\text{SC}}\mathcal{L}^{\text{DC}}\mathcal{L}^{\text{T}} \tag{5.2}$$

$$= \prod_t \prod_j^m \mathcal{L}^t_{\text{Eff}}(\mu_j(\vec{\theta}, \vec{\xi}), \sigma_j(\vec{\theta}, \vec{\xi}); n_j), \tag{5.3}$$

with single cascade, double cascade and track topologies $t = \text{SC,DC,T}$. $\vec{\theta}$ are the parameters of interest, $\vec{\xi}$ the nuisance parameters, μ_j the expected number of events in the j-th bin with statistical uncertainty σ_j, and n_j the observed number of events in the j-th bin. It is an extension of the likelihood used for the diffuse fits assuming a neutrino flavor composition of $1:1:1$ by the two parameters of interest describing the relative flavor contributions. The SAY likelkihood is an effective likelihood, using the effective number of MC events $n_{\text{eff}} = (\sum_i w_i)^2 / \sum_i w_i^2$ and the effective weight $w_{\text{eff}} = \sum_i w_i^2 / \sum_i w_i$ to obtain the expected number of events, $\mu = w_{\text{eff}}\, n_{\text{eff}}$.
The flavor composition parameter space is scanned in order to obtain a two-dimensional confidence region for the flavor composition, which can then be compared to the expected flavor composition for astrophysical neutrinos shown in Figure 2.3. The confidence region is obtained assuming Wilks' theorem [151] which states: If the parameters of interest and the nuisance parameters are not bounded, and for a large number of observed events n, the likelihood ratio $-2\Delta \log \mathcal{L} = -2 \log(\mathcal{L}_{\text{test}}/\mathcal{L}_{\text{max}})$ for nested models is χ^2_k-distributed,

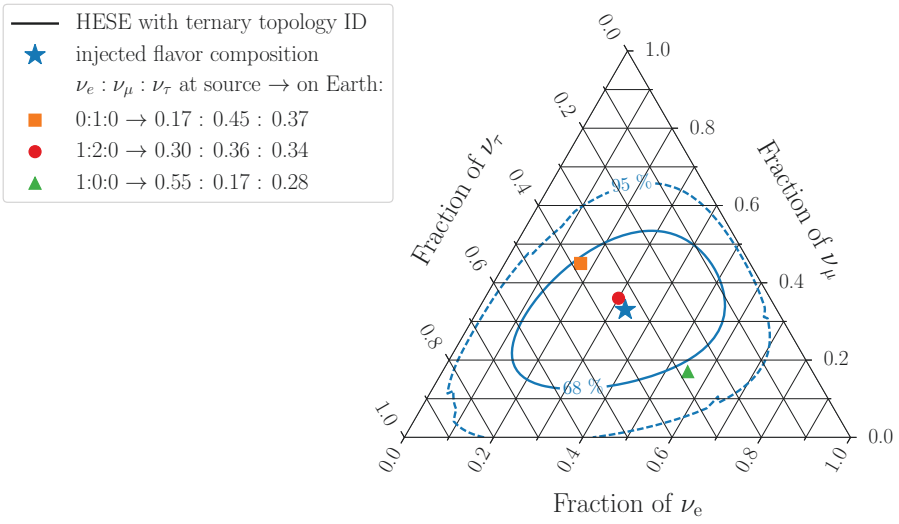

Figure 5.7: Asimov sensitivity to the flavor composition measurement, for an astrophysical neutrino spectrum following the single unbroken power-law given in Equation 5.1, with a flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$, and the baseline atmospheric spectra described in Section 5.2. Contours show the 1σ and 2σ confidence intervals assuming Wilks' theorem holds.

where k is the difference in the number of degrees of freedom of the tested scenario with likelihood $\mathcal{L}_{\text{test}}$ to the best fit with maximum likelihood $\mathcal{L}_{\text{max}}$. In this case, the tested scenarios are the tested flavor compositions, and, using $\vec{f}_\alpha = \nu_e : \nu_\mu : \nu_\tau$, we obtain

$$\text{TS} = -2\Delta \log \mathcal{L} = -2 \log \left(\frac{\mathcal{L}(\vec{n}_j \mid \vec{\mu}(\vec{f}_\alpha^{\,\text{t}}, \hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\xi}}}))}{\mathcal{L}(\vec{n}_j \mid \vec{\mu}(\hat{\vec{f}}_\alpha, \hat{\vec{\theta}}, \hat{\vec{\xi}}))} \right) \tag{5.4}$$

where $\hat{\vec{f}}_\alpha, \hat{\vec{\theta}}, \hat{\vec{\xi}}$ are the fit parameters maximizing the likelihood, and $\hat{\hat{\vec{\theta}}}$ and $\hat{\hat{\vec{\xi}}}$ are the fit parameters maximizing the likelihood under the constraints of the tested scenario $\vec{f}_\alpha^{\,\text{t}}$. $k = 2$ due to the constraint that the fractions for each flavor have to sum up to one, $\sum_{\alpha=e,\mu,\tau} f_\alpha = 1$.

Using this prescription for the confidence region, the median expected sensitivity is shown in Figure 5.7 under the assumption of a single unbroken power-law with $\Phi_{\text{astro}} = 6.9$, $\gamma = 2.5$ and $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$. It is generated using the Asimov dataset [152], containing all possible realizations of the experiment. The contours show the 1σ and 2σ confidence interval assuming Wilks theorem holds, even if the sample size is small and the parameters are bounded. Since a $\chi^2_{k=2}$ is assumed, the 1σ region is given by $-2\Delta \log \mathcal{L} \leq 2.3$, and the 2σ region is given by $2.3 < -2\Delta \log \mathcal{L} \leq 6.2$. It can be seen

that with an expected 4 events classified as double cascades, the statistical uncertainties are large. Therefore, a non-zero ν_τ flux can only barely be disfavored at 2σ in the Asimov case. It has to be noted that the previous update of HESE with 6 years of livetime fit a spectral index of $\gamma_{\text{HESE-6}} = 2.92$ [132], much softer than the multi-sample analysis [32]. With a softer spectrum, fewer events are expected, both in total, and in particular with a double cascade classification. Further, the ν_τ-induced signal in the double cascade topology class also decreases with a softer spectrum, as fewer of the ν_τ-CC events in the HESE sample will be at high energies with $L_\tau > 10$ m.

5.4 Influence of Systematic Uncertainties

The 7.5-year HESE analyses extend the treatment of systematic uncertainties as compared to the previous iterations. Two classes of systematic uncertainties are considered: **modeling uncertainties** leading to a limited or biased knowledge of the atmospheric backgrounds, and **detector systematics** affecting the inferred observables of the events. The atmospheric spectrum is weighted assuming the Hillas-Gaisser three-population model (H3a) for the incident CR spectrum [139]. Uncertainties in the conventional and prompt atmospheric normalization scales are included as parameters Φ_{conv} and Φ_{prompt}, respectively. For the shower development in the atmosphere the hadronic model **Sibyll 2.3c** [140] is used. Both of these models must be treated with care. The influence of the incident cosmic-ray spectrum and the choice of the hadronic model was investigated, and found to have a negligible influence on the measurement of the astrophysical flux parameters [136]. The former is taken into account in the fitting as $\Delta\gamma_{\text{CR}}$, i.e. the deviation of the cosmic ray spectral index from the nominally assumed one. For the hadronic model, the parameter $R_{K\pi}$ describing the ratio of Kaons to pions in the shower. As the cross-sections are still different for neutrinos and antineutrinos at the lower end of the HESE fitting range of 60 TeV, and IceCube cannot tell them apart, variations the neutrino-to-antineutrino ratio, $2\nu/(\nu + \bar{\nu})_{\text{atm}}$ are considered. Further, modeling of the CR-induced shower affects the normalization of the muon flux Φ_μ, which is derived from data: a second veto layer, inside of the first one, is defined, and events are counted that pass the initial veto layer but fail the second one. Accounting for the different effective volumes of these layers, the muon normalization can be obtained as

$$\Phi_\mu = 2.1 \frac{\Phi_{\text{MC}}}{N_{\text{MC}}} \times N_{\text{tagged}}, \tag{5.5}$$

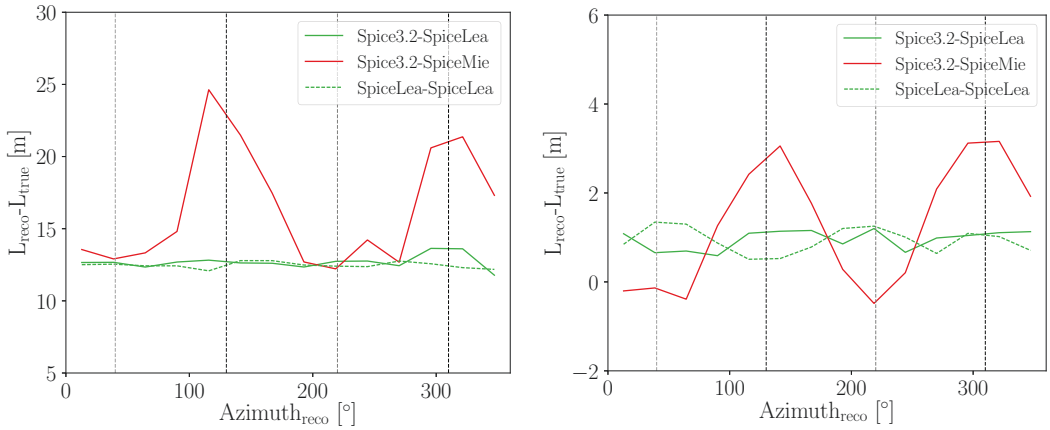

Figure 5.8: Length bias for true single cascades (*left*) and true double cascades (*center*) classified as double cascades. The length bias is shown for events simulated with Spice3.2 and reconstructed with SpiceLea effective distance spline tables (green solid), events simulated with Spice3.2 and reconstructed with isotropic SpiceMie spline tables (red), and for comparison events simulated with SpiceLea and reconstructed with SpiceLea effective distance spline tables as used in the analysis presented in [106].

where the factor 2.1 describes the volume normalization, Φ_{MC} and N_{MC} are the flux and number of muons, respectively, as obtained from simulation and N_{tagged} is the number of tagged muons that deposit light in the first but not in the second layer of the. For details, see [91].

On the detector side, one parameter affecting the energy inference is the total DOM efficiency to convert photons into current, included in the fit as ϵ_{DOM}. The influence of the DOM efficiency is only on the total energy scale, i.e. it does not affect the spectral index measurement or the flavor ratio, however, it affects the total normalization which is given at 100 TeV. The diameter of the bubble column in the hole ice affects the angular acceptance of a DOM and is included as $\epsilon_{\text{head-on}}$. Further, the global scattering and absorption parameters are taken into account by testing the changes to the analyses when the underlying MC sets are swapped for one of the sets with varied bulk ice parameters. All these systematic uncertainties do not affect the flavor composition measurement in a significant way. The atmospheric modeling might affect the atmospheric backgrounds, predominantly from atmospheric ν_μ, however, the major backgrounds for the identification of ν_τ-induced double cascades are events caused by astrophysical ν_e and to somewhat lesser extent by astrophysical ν_μ. Uncertainties on the neutrino-nucleus cross-section are not considered, as they are degenerate with the measurement of the astrophysical flux normalization.

The major systematic uncertainty for the identification of ν_τ-induced double cascades is the ice anisotropy introduced in Section 3.3, which can cause biases in the reconstructed double cascade length L_{dc}. As the photon scattering is inhibited along the major anisotropy axis by 10% (see Figure 3.3), the photons propagate further and faster. If a single cascade is aligned with the anisotropy axis, it will become elongated and may mimic a double cascade, as the timing of the light arriving at the DOMs is also changed due to less scattering. On the contrary, perpendicular to the major anisotropy axis, scattering is enhanced by 5%, leading to photons propagating less far, and slower. A double cascade aligned with one of the minor anisotropy axes will be compressed and thus may look like a single cascade. If not accounted for, the anisotropy will lead to a higher misclassification of true single cascades as double cascades along the major anisotropy axis, and to a lower classification efficiency of true double cascades along the minor anisotropy axes. However, the treatment of ice anisotropy in reconstruction is very difficult, and cannot be done precisely. The reconstruction uses the photo spline tables introduced in Section 3.4.3. They contain tabulated light yields for simulated 1 GeV cascades, which are placed $\Delta z = 20$ m apart in the center of the detector in x, y and between depths of -600 to 600 m, and have zenith angles with $\Delta\theta = 10^\circ$ ranging from 0° to 180°. The light propagation is assumed to only depend on the ice layer the light is propagating through, i.e. it is a function of the depth and the zenith angle. The anisotropy would require to add the azimuth as additional dimension, and the ice layer tilt leads to the light yield also being a function of x and y, for a total of three additional dimensions. Due to finite computational resources, it is not possible to extend the tables in such a way. A breakthrough in the treatment of the anisotropy was achieved by introducing effective distance spline tables, in which the light yield is looked up not at the actual position of the DOM, but at the isotropic-ice-equivalent position. The development and performance of the effective distance spline tables are discussed at length in [106]. The spline tables are produced for a specific ice model, in this case for the SpiceLea model. Since their development, the ice model was updated to Spice3.2, with slightly different values for the scattering and absorption coefficients, and a higher value for the anisotropy. While the new ice model was used in simulation production, no new spline tables for reconstruction were produced. Naturally, the question arises, if the SpiceLea effective distance spline tables can be applied to Spice3.2 simulation, and if new biases are introduced. Figure 5.8 shows the effectiveness of the anisotropy correction by using the SpiceLea effective distance spline tables. It can be seen that the bias is corrected just as well for events simulated with the newer Spice3.2 ice model, as with the older SpiceLea ice model, for which the tables were originally developed. Note

that single cascades have true lengths of 0 m. The seemingly high length bias for true single cascades in the double cascade subsample stems from the fact that only single cascades with a reconstructed $L_{\mathrm{dc}} \geq 10$ m can be classified as double cascades, making for a minimal length bias of 10 m.

The uncertainty in the absolute magnitude of the variation of the scattering parameters due to the anisotropy along the major anisotropy axis is described by the parameter a_s, the total scale of the anisotropy. An uncertainty in the directions of the anisotropy axes is not modeled – it is assumed to be much smaller than the angular uncertainty of single cascades.

Chapter 6

(Re-)Opening the Box

Previous to this analysis, 6 years of data have been analyzed using the HESE selection. The analyses were performed on *pass-1* data assuming isotropic phhoton propagation through the ice, and included an analysis of the diffuse spectrum and a search for neutrino sources [132]. The 6-year HESE sample was then used for a measurement of the flavor composition using the ternary topology classification, which included a consistent processing of all events using effective distance spline tables. No events with energies $E_{\text{tot}} \geq 60$ TeV were classified as double cascades [37].

The HESE sample was (re-)unblinded in May 2018 for this analysis and the analyses introduced in Section 4.3. It contains 102 events passing the HESE selection, i.e. the veto and total charge criteria described in Section 4.2. It has a livetime of 2635 days, approximately 7.5 years, and contains data from 2010 to 2017. The initial six years of livetime were re-unblinded with a complete re-processing of the events from tape following the improved detector calibration (*pass-2*). This led to the removal of eight events from the sample, as they now fail one of the selection cuts. The new detector calibration led to an overall decrease of the number of PE associated with an event, leading to an overall decrease of the reconstructed energies. Both the total charge and the veto selection criteria include the counting of PE of an event. Thus, a decrease of an event's total charge can cause it to fail the requirement of a minimum of 6000 PE total charge or the veto cut allowing a maximum of 3 out of the first 250 PE deposited in the veto region. Seven of the removed events now fail the total PE criterion, out of which three events previously had a reconstructed energy of $E_{\text{tot}} > 60$ TeV; one track now fails the veto requirement. Further, one event was on a tape that could not be read anymore, and is therefore lost to the sample now.

The re-processing also affects the reconstructed total deposited energy, E_{tot}. As only

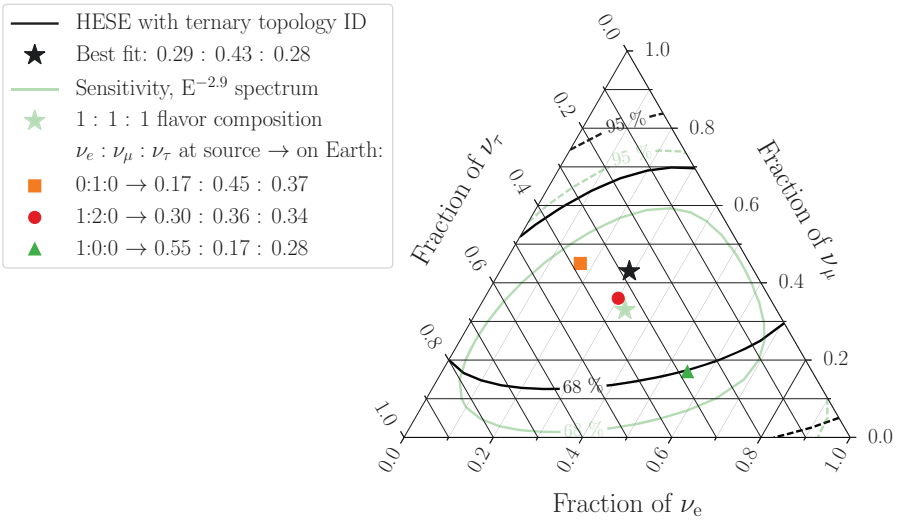

Figure 6.1: Results of the flavor composition measurement (black lines) and sensitivity at the best-fit single-power-law spectrum with a fixed 1 : 1 : 1 flavor composition (light green lines). Contours show the 1σ and 2σ confidence intervals assuming Wilks' theorem. Flavor compositions expected from popular source scenarios are marked.

events with $E_{\text{tot}} \geq 60$ TeV are used in the fits, some events pass the HESE selection cuts, but do not contribute to the fits, and are not classified into the topologies. Three events in the HESE sample that were previously reconstructed to > 60 TeV, now fall under the 60 TeV threshold. There are 60 events with $E_{\text{tot}} > 60$ TeV in the sample with 7.5 years of livetime, with 17 of these observed in the newly added ~ 1.5 years of data: 12 events were observed in the 2016 season, and 5 events were observed in the first half of the 2017 season. Based on these 60 events, the flavor composition analysis is performed and presented in Section 6.1. Two events are classified as double cascades, their observables are discussed in Section 6.2. A Discussion of systematic uncertainties in Section 6.3 concludes this chapter.

6.1 Results of the Flavor Composition Analysis

The 60 events with reconstructed total deposited energy of $E_{\text{tot}} \geq 60$ TeV are classified into 42 single cascades, 2 double cascades and 16 tracks. Those are the first events classified as double cascades in IceCube. The single power-law fit with a flavor composition

True Topology	Reconstructed Topology single	double	track	Reconstructed Topology single	double	track
single	68%	26%	11%	68%	28%	10%
double	23%	63%	6%	21%	59%	5%
track	9%	11%	83%	11%	14%	86%
expected events	43.2	2.3	13.5	42.2	2.2	15.3
observed events	42	2	16	42	2	16
$\nu_e : \nu_\mu : \nu_\tau$	1:1:1			0.29:0.43:0.28		

Table 6.1: True topology vs. each reconstructed topology for the best-fit astrophysical and atmospheric spectra given in the text. The columns sum up to 100%. The entries on the diagonal show the purity of each reconstructed topology, while the off-axis entries show the misidentification fraction. The last lines give the total number of expected events classified into each topology with the best-fit spectrum, and the actual observed number of events per topology.

fixed to $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$ yields:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}E} = 6.37 \cdot \left(\frac{E}{100\ \mathrm{TeV}}\right)^{-2.87} \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1} \tag{6.1}$$

as best fit, with $\Phi_{\mathrm{astro}} = 6.37^{+1.5}_{-1.6}$ and $\gamma_{\mathrm{astro}} = 2.87^{+0.20}_{-0.19}$. The atmospheric parameters are fit to $\Phi_{\mathrm{prompt}} = 0^{+5.3}$, $\Phi_{\mathrm{conv}} = 1.01^{+0.35}_{-0.33}$, and $\Phi_\mu = 1.19^{+0.45}_{-0.44}$. The expectation for these spectra is to observe 59.0 events, classified into 43.2 single cascades, 2.3 double cascades (1.5 ν_τ-induced, 0.8 background) and 13.5 tracks. The single power-law fit with free flavor contributions yields:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}E} = 6.98 \cdot \left(\frac{E}{100\ \mathrm{TeV}}\right)^{-2.89} \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1} \tag{6.2}$$

as best fit, with $\Phi_{\mathrm{astro}} = 6.98^{+2.4}_{-2.1}$, $\gamma_{\mathrm{astro}} = 2.89^{+0.21}_{-0.20}$, and $\nu_e : \nu_\mu : \nu_\tau = 0.29 : 0.43 : 0.28$. The atmospheric parameters are fit to $\Phi_{\mathrm{prompt}} = 0^{+6.1}$, $\Phi_{\mathrm{conv}} = 0.97^{+0.36}_{-0.36}$, and $\Phi_\mu = 1.16^{+0.46}_{-0.45}$. The expectation for these spectrum is to observe 59.8 events, classified into 42.2 single cascades, 2.2 double cascades (1.4 ν_τ-induced, 0.8 background) and 15.3 tracks.

The results of a 2D-likelihood scan of the flavor contributions are shown in Figure 6.1. As can be seen, the contours are quite wide and do not disfavor a vanishing ν_τ-flux. The ν_τ-fraction is strongly dependent on the two-dimensional MC PDFs for the double cascade topology shown in Figure 6.2, while the classification of more events as tracks than expected from a $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$ composition drives the best-fit to a higher

	Reconstructed Topology			Reconstructed Topology		
Flavor	single	double	track	single	double	track
ν_e	57%	21%	10 %	56%	22%	8%
ν_τ	28%	66%	11%	26%	62%	8%
ν_μ	15%	13%	73%	18%	17%	78%
μ_{atm}	0%	0%	7%	0%	0%	6%
$\nu_e : \nu_\mu : \nu_\tau$	1:1:1			0.29:0.43:0.28		

Table 6.2: Flavor content for each reconstructed topology for the best-fit astrophysical and atmospheric spectra given in the text. The columns sum up to 100%.

ν_μ-fraction. The width of the contour also depends strongly on the astrophysical spectral index, which is softer than the baseline spectrum assumed.
Table 6.1 shows a comparison of the true topology content in each reconstructed topology subsample for the best fits assuming a single-power-law spectrum and a fixed flavor composition with $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$, and for the single-power-law with unconstrained flavor composition. Note that true double cascades include also ν_τ-CC induced events, in which the tau decays after less than 10 m. The total number of expected events classified into each topology with the best-fit spectrum, and the actual observed number of events per topology are also shown. The observed number of events classified into each topology matches the number of events per topology expected from the flavor composition fit.
Table 6.2 shows the flavor content of each reconstructed topology sub-sample, again for the fixed-flavor-composition and free-flavor-composition best fits. The mapping between reconstructed topology and underlying neutrino flavor is less straight-forward. While only ν_τ interactions can create (true) double cascades, both ν_μ and ν_τ can create (true) tracks. All flavors can create (true) single cascades via the NC interaction, and additionally many ν_τ-CC interaction will produce a reconstructed single cascade, as the double cascade length will be too short to be resolvable with currently available techniques.
In the following, the reconstructed properties of the two double cascades are discussed, particularly focussing on whether these are more signal-like, i.e. ν_τ-induced, or background-like.

6.2 Reconstructed Properties of the Two Double Cascades

The flavor composition measurement method relies on the contributions of the different components, both astrophysical and atmospheric, scaling up and down. This is done by comparing two-dimensional MC templates with the observed data. Figure 6.2 shows

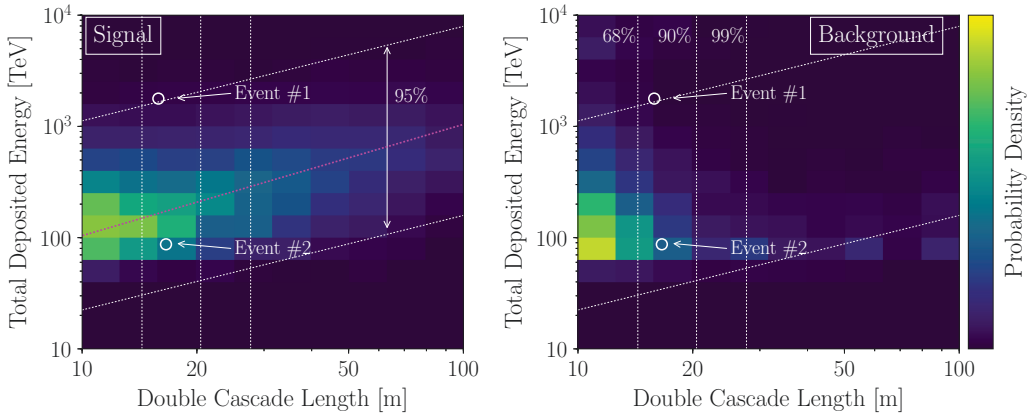

Figure 6.2: Two-dimensional PID PDFs showing total reconstructed energy versus reconstructed double cascade length for the double cascade sample with data points, using the best fit to the atmospheric and astrophysical components with the flavor composition fixed to 1 : 1 : 1. In both the signal (ν_τ-induced double cascade events) histogram (*left*),and the background (all remaining events) histogram (*right*), the two tau-neutrino candidates are overlaid as white circles. See text for details.

the two-dimensional MC PDF templates for the double cascade topology, at the best-fit spectral parameters given in Equation 6.1. The left panel shows the distribution of ν_τ-induced double cascades, which are the "signal" component of the double cascade topology. The median value of the distribution is indicated by the magenta line. The right panel shows the ν_e and ν_μ-induced events that are misclassified as double cascades, these constitute the "background" contribution to the double cascade topology. The two events classified as double cascades are overlaid as white circles. The vertical lines on the right panel mark how fast the single-cascade background drops as a function of length, i.e. 68% of misclassified single cascades have reconstructed double-cascade lengths of < 14.4 m, 90% have double-cascade lengths below 20.4 m and only 1% has double-cascade lengths above 27.5 m. Both events are at higher L_{dc} than 68% of the single-cascade background. It can be seen that event #1 lies in a region of the parameter space where little signal and little background is expected, it furthermore lies on the very edge of the region containing 95% of ν_τ-induced double cascades. This is due to its high reconstructed energy and the short double cascade length for that energy. Event #2 on the other hand lies in a region in the parameter space which is occupied by both signal and background.

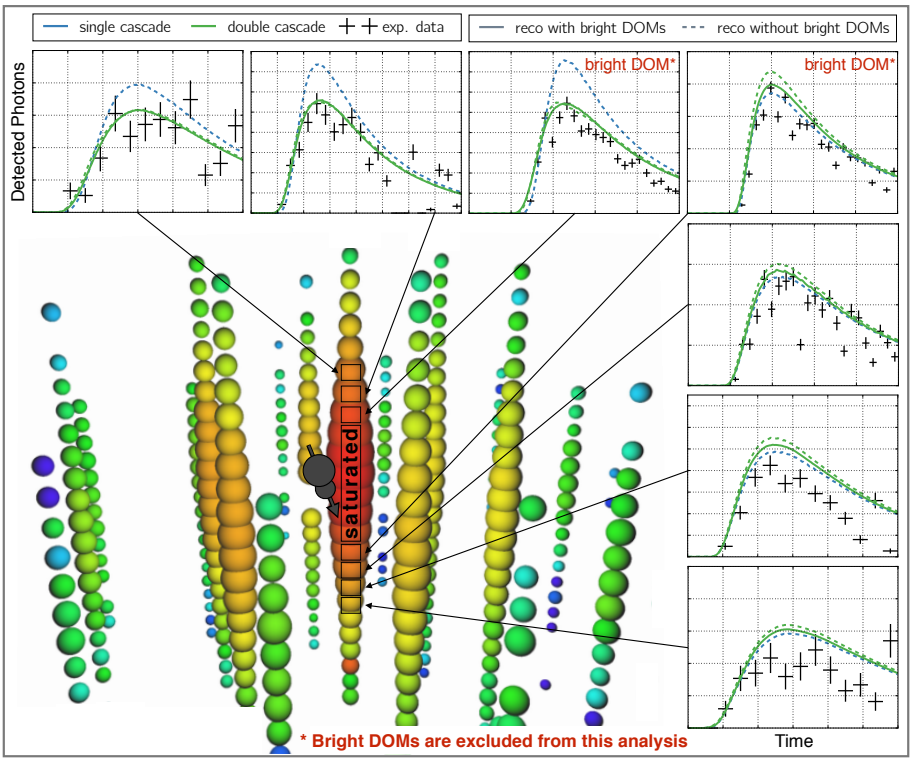

Figure 6.3: Event view of double cascade event #1: "Big Bird". The reconstructed double cascade vertices are indicated as grey circles, the direction indicated with a grey arrow. The saturated DOMs close to the double cascade vertices are labelled. The minor panels show the observed PE count rates over time along with the predicted PE PDFs for a single cascade and double cascade hypothesis.

Event #1: "Big Bird"

Event #1 was observed in 2012, but had so far been classified as a single cascade. It is the highest-energy event in the HESE sample and was nicknamed "Big Bird". The event view obtained with the `Steamshovel` visualization program [153] is shown in the main panel in Figure 6.3. The grey circles mark the reconstructed cascade positions, with the arrow indicating the reconstructed direction. The size of the circles illustrates the relative deposited energy of the two cascades. In the minor panels, the photon counts over time are displayed alongside with the predicted photon count PDFs for a single cascade and double cascade hypothesis for several DOMs. The photon count rates do not show a clear preference for a single vs. a double cascade hypothesis. Note that the two labeled bright DOMs (see Section 3.4.3) were excluded from the reconstruction. Further, six DOMs near the interaction vertex are saturated and excluded from the reconstruction.

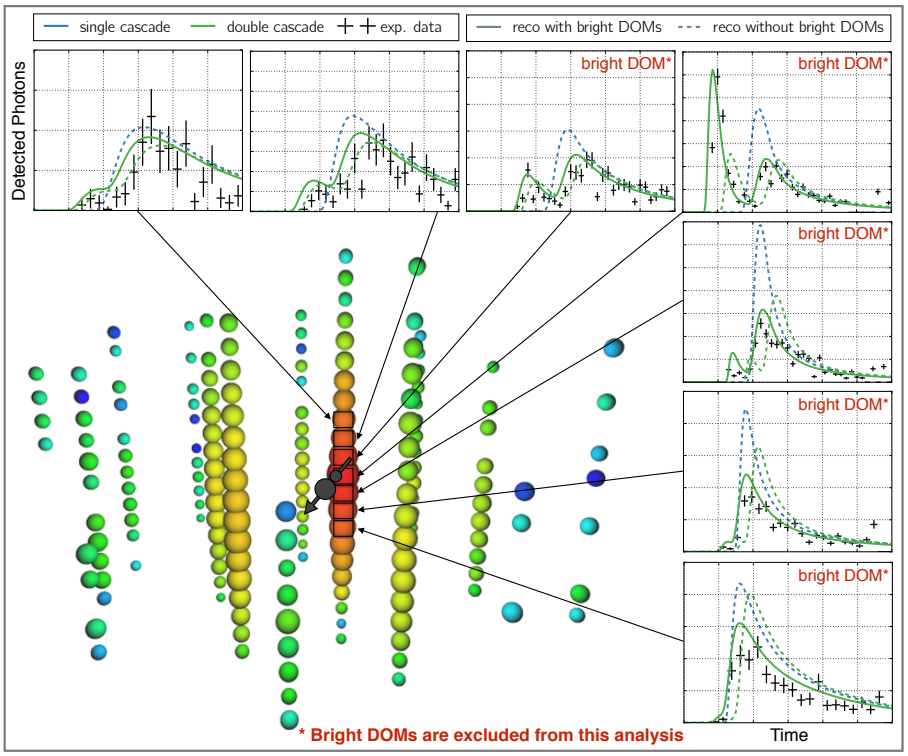

Figure 6.4: Event view of double cascade event #2: "Double Double". See Figure 6.3 for details.

Thus, "Big Bird" is reconstructed without the eight DOMs closest to its reconstructed vertex.

Event #2: "Double Double"

Event #2 was observed in 2014, but had so far been classified as a single cascade. The event view is shown in the main panel of Figure 6.4, while the photon counts over time are displayed alongside with the predicted photon count PDFs for a single cascade and double cascade hypothesis for several DOMs in the minor panels. As can be easily seen, the predicted photon count PDFs differ remarkably between the single and double cascade hypothesis, with the single cascade hypothesis disfavored according to the observed counts. The event is very indicative of two separate energy depositions in the ice. Therefore it was nicknamed "Double Double". "Double Double" has five bright DOMs that were ignored in the reconstruction.

The reconstructed properties of both events classified as double cascades are shown in Table 6.3. Comparing these properties with the classification scheme shown in Figure

	"Big Bird"	"Double Double"
Energy of 1st cascade	1.2 PeV	9 TeV
Energy of 2nd cascade	0.6 PeV	80 TeV
Energy Asymmetry	0.29	-0.80
Length	16 m	17 m
HESE-6 classification	single cascade	single cascade

Table 6.3: Reconstructed properties of the two Double Cascades.

5.2, the following conclusions can immediately be drawn:

- Both double cascades have $L_\tau > 15$ m, which is larger than what most true single cascades in the double cascade sample are reconstructed to.
- "Big Bird" has a small double-cascade length L_τ given its high deposited energy, lower than what is expected for a ν_τ-interaction, see also Figure 6.2.
- "Big Bird"'s energy asymmetry is very close to the the edge of the selected interval at 0.3, which was chosen assuming a harder spectrum to reduce single cascade contamination.
- "Double Double"'s energy asymmetry is -0.8, where background contamination is low.

The distribution of the reconstructed energy asymmetry for simulated events in the double cascade sample is shown in Figure 6.5 for the best-fit astrophysical and atmospheric spectra and a flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$. The data events are shown in black. It can be seen that "Double Double" at $A_E = -0.80$ is in a signal dominated region, while "Big Bird" at $A_E = 0.29$ lies in a region where the background contributions dominate. It can also be seen that the energy asymmetry distribution is flat for correctly classified double cascades, but not for misclassified single cascades. The single cascade contamination in the double cascade topology sub-sample is small at moderate negative A_E, but increases towards the two edges of the selected range at $A_E = -0.98, 0.30$.

How likely is the observation of ν_τ-CC events with such properties?
In principle, one can use $\langle L_\tau \rangle \sim 50 \text{ m} \cdot E_\tau/\text{PeV}$ to estimate how likely it is to observe taus with the decay lengths and decay cascades' energies as the found double cascades. However, $\langle L_\tau \rangle$ can have large event-by-event fluctuations.
A better gauge if observations match expectations is provided when looking at the distribution of the ratio of the double cascade length to the decay cascade energy in the

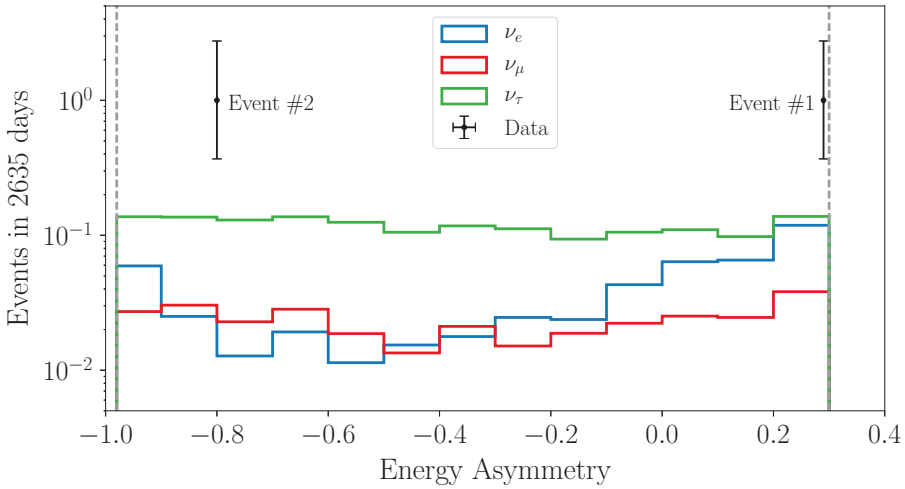

Figure 6.5: Distribution of simulated reconstructed energy asymmetry in the double cascade subsample split by underlying flavor for the best-fit astrophysical and atmospheric spectra. The values of the two double cascades are shown.

double cascade subsample, which is shown in Figure 6.6. It includes the selection bias towards larger L_τ. "Double Double" lies very close to the median value expected for selected ν_τ-induced double cascades, while "Big Bird" lies outside the 90% but within 95% of the distribution. Both events lie in a signal dominated region.

6.3 Systematic Uncertainties on the Classification

As discussed in Section 5.4, the anisotropy in the photon scattering parameters in the ice can influence double cascade reconstruction and identification. In particular, events aligned with the major anisotropy axis along 130° and 310° in the azimuthal direction tend to get reconstructed to larger L_{dc}. If uncorrected, this effect leads to a higher contamination of the double cascade subsample by true single cascades. In this analysis, the effect of the anisotropy is corrected for by using the effective spline tables developed in [106], but given that both events classified as double cascades were previously classified as single cascades with the same method, the possible influence of the ice anisotropy is the first systematic to check. The minor anisotropy axes are perpendicular to the major axis, at 40° and 220° in azimuth and aligned with the $z-$axis. The reconstructed azimuthal angles for "Big Bird" and "Double Double" are 348° and 9°, respectively. Both events are downgoing, with reconstructed zenith angles of 32° and 48°, respectively. Thus, none of the events are aligned with any of the anisotropy axes. We conclude that it is unlikely for either event's topology classification outcome to be influenced by the

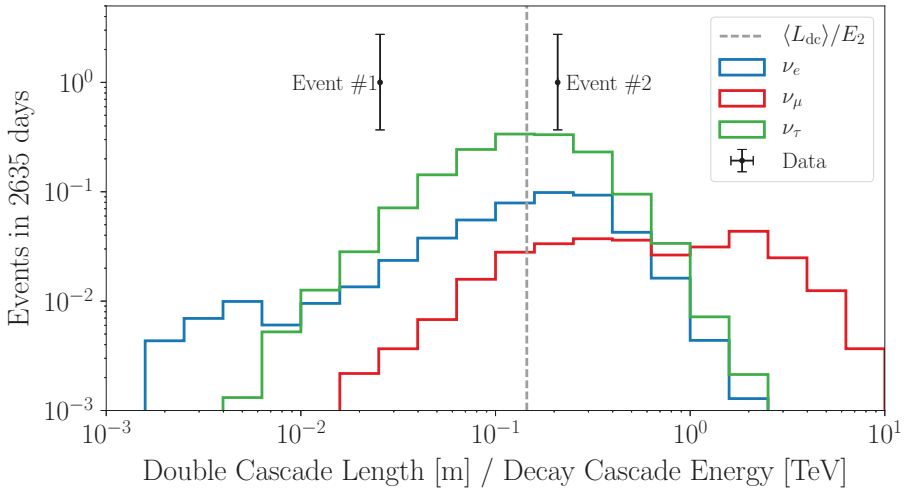

Figure 6.6: Distribution of the ratio of double cascade length to reconstructed decay cascade energy in the double cascade subsample split by true topology for the best-fit astrophysical and atmospheric spectra based on simulation. The values of the two double cascades and the median value for ν_τ-induced double cascades are shown.

ice anisotropy.

Why were the double cascades classified as single cascades before?

The next question is how much the new detector calibration and data reprocessing affects the classification. Previously, the events were processed using the initial, *pass-1* calibration with an overall detector-averaged correction of the SPE peak, while the new *pass-2* detector calibration uses the proper SPE peaks measured for each individual DOM. Table 6.4 shows the resulting classification of "Big Bird" using the old, *pass-1*, and the new, *pass-2*, detector calibration. Indeed, the previous single-cascade classification of "Big Bird" can be recovered with every combination but the new processing used in this analysis. Further, "Big Bird" is classified as a single cascade when bright DOMs are included. The old [106] and new official processing and resulting classifications are highlighted in gray. This comparison is repeated for "Double Double" and the results shown in Table 6.5. Processing *pass-2* data needed some changes in the software. It can be seen that all reconstructions performed using the newer software version consistently result in a double cascade classification of "Double Double", while reconstructions performed using the older, *pass-2*-incompatible software version[1] yield

[1]The combination of *pass-2* processing and *pass-2*-incompatible software required a change of the pulsemap structure. This was only performed for this particular event for comparison reasons.

software	new	new	new	old	old
calibration	pass-2	pass-1	pass-2	pass-1	pass-1
bright DOMs	excl.	excl.	incl.	excl.	incl.
topology	d.c.	s.c.	s.c.	s.c.	s.c.

Table 6.4: Topology classification of "Big Bird" when using different software and data processing combinations. *Pass-2* processing follows the new detector calibration as used in HESE-7, *pass-1* follows the initial detector calibration; "new" software refers to the software version used in HESE-7, while "old" software contains the minimizer tolerance bug (see text). By default, bright DOMs are excluded in high-energy analyses, the influence of their inclusion on the final topology classification (d.c. for double cascade, s.c. for single cascade) is also shown. The first highlighted column is the result presented in this work, the second highlighted column refers to the previous result obtained on the 6-year sample [106].

software	new	new	new	new	old	old	old	old
calibration	pass-2	pass-1	pass-2	pass-1	pass-2	pass-1	pass-2	pass-1
bright DOMs	excl.	excl.	incl.	incl.	excl.	excl.	incl.	incl.
topology	d.c.	d.c.	d.c.	d.c.	s.c.	s.c.	d.c.	d.c.

Table 6.5: Topology classification of "Double Double" when using different software and data processing combinations. See caption of Table 6.4 for further details.

a double cascade classification only if bright DOMs are included. It turned out that a bug in the minimizer setting has been fixed in the newer software version, decreasing the minimizer tolerance by a factor 100. This led to very small changes in MC processing that were not noticeable in the performed MC comparisons. However, it seems that the likelihood landscape is not as smooth for this data event as it is for the majority of MC simulated events. The too high minimizer tolerance then led to not finding the narrow global minimum in the previous analysis. In fact, with the old software version and the fixed minimizer bug, the double cascade classification of "Double Double" is reproduced. The flat likelihood landscape is probably exacerbated by the removal of the bright DOMs, which show the largest discrepancy between single-cascade expectation and recorded waveform (see Figure 6.4). As explained in Section 3.4.3, bright DOMs can dominate the likelihood and bias the reconstruction of high-energy events. "Double Double" has a reconstructed energy of $E_{\text{tot}} = 89$ TeV, which is an order of magnitude lower than the PeV events that the removal of the bright DOMs was introduced for. The question arises, if the blanket removal of bright DOMs for all HESE events is a good choice. Reconstructing "Double Double" with bright DOMs included yields a double

cascade classification for both *pass-1* and *pass-2* data processing. Even with the minimizer bug, the double cascade classification is recovered if bright DOMs are included. This fact as well as comparable reconstructed properties when the bright DOMs are included strengthens the confidence in the proper classification of "Double Double" as a double cascade. "Big Bird" on the other hand is reconstructed as a single cascade once bright DOMs are included. However, as mentioned before, the removal of bright DOMs was introduced precisely for very-high-energy events such as "Big Bird". In the next iteration of a flavor analysis using `taupede` or the next medium-and-high-energy analysis, the influence of bright DOMs should be carefully assessed, and perhaps a better prescription found for their inclusion or removal.
Finally, the different software and data processing combinations shown in Tables 6.4 and 6.5 also yielded slightly different `monopod` reconstructions which serve as seed vertices and directions for `taupede`. For "Double Double", also the influence of the `taupede` seed directions on the classification outcome was studied. This study is presented in Appendix A.

Chapter 7

A Posteriori Analysis of Tau Neutrino Candidate Events

In the previous chapter, the double cascades were presented as tau neutrino candidates and their reconstructed properties were studied in context of the original analysis. The conclusion was that "Double Double" shows characteristics expected from a ν_τ-CC interaction, while "Big Bird" shows characteristics consistent with being a misclassified single cascade. Unfortunately, the original analysis and the produced MC statistics do not allow a proper quantitative assessment of how tau-like, or how background-compatible, each of the double cascades is. But by generating a large sample of MC events with similar properties as the data events in a dedicated resimulation a quantitative statement can be made. This will be discussed in this chapter, starting with the description of the resimulation setup in Section 7.1. In Section 7.2, a method for a proper statistical treatment of the events on a multi-dimensional parameter space is introduced. The updated flavor composition measurement is presented and discussed in Section 7.3 and to conclude this chapter, source flavor composition scenarios are discussed in Section 7.4.

7.1 Resimulation of Tau Neutrino Candidate Events

The flavor composition measurement relies on a binned likelihood analysis using only the two dimensions L_{dc} and E_{tot} for the double cascade topology as shown in Figure 5.6, and $\cos(\theta_z)$ and E_{tot} for the single cascade and track topologies in Figures 5.4 and 5.5, respectively. Other properties are used to assign the final topology classification, as shown in Figure 5.2. However, they are not used to evaluate how tau-like an event

with a double cascade classification is. As seen in Figure 6.5, the energy asymmetry shows a rather flat distribution for true double cascades, but not for misclassified true single cascades. Obviously the flavor composition measurement would gain sensitivity, if the distribution of the energy asymmetry was taken into account. However, the existing MC statistics are too small to extend the likelihood analysis to a third dimension. It is also unfeasible to produce the MC statistics necessary for such an extension, as that would require to increase the statistics by more than one order of magnitude, and computational resources are limited. Therefore, new, targeted MC sets were generated specifically resimulating events similar to the observed ones.
Resimulating events in IceCube is a standard procedure to assess a rare event's compatibility with background, and has been performed since the earliest days of HESE to assess the reconstruction resolution and the probability of atmospheric origin of the found high-energy events that were claimed to be astrophysical neutrinos. The resimulation uses the reconstructed properties of an event, like the reconstructed energy, direction, and vertex, to generate MC events with similar true properties. A starting track would typically be resimulated by generating ν_μ-CC events and atmospheric muons, while for a high-energy single cascade ν_e-CC and perhaps ν_e-NC would be simulated. For the two double cascades, the situation is slightly more complicated:

- All background scenarios have to be tested, i.e. a resimulation of neutrinos of all flavors is necessary, as well as atmospheric muons for "Double Double".
- As shown in Section 6.3, using different software versions and reconstruction settings can affect the reconstructed properties of the events. In order to avoid biasing the results, several versions of the event properties should be included in the resimulation.
- The mapping between true neutrino energy and reconstructed deposited energy is only simple for ν_e-CC interactions, as discussed in Section 4.1.
- True values for L_τ only exist for ν_τ-CC interactions.

The restricted parameter space of the resimulation is shown in Table 7.1. The mapping between true and reconstructed quantities is not straightforward. For the interaction vertex, the resimulation range was chosen to lie in a cylinder with radius and height of 50 m, centered on the reconstructed vertex of the observed events. This range is large compared to the ~ 2 m localization resolution of the shower maximum and the displacement of the shower maximum from the interaction vertex of not more than a few meters. For the zenith and azimuth angles, the resolution depends on the event topology. The

Variable	"Big Bird"	"Double Double"
Primary Energy E_ν	> 65 TeV	> 1.5 PeV
Visible Energy E_{vis}	60 - 300 TeV	1 - 3 PeV
Vertex, $r - r_{\text{evt}}$	50 m	
Vertex, $z - z_{\text{evt}}$	25 m	
Azimuth $\phi - \phi_{\text{evt}}$	$\pm 110°$	
Zenith $\theta - \theta_{\text{evt}}$	$\pm 35°$	

Table 7.1: Restricted parameter space for the resimulation of the double cascades. The upper value of the primary energy depends on the interaction type, reflecting the spread of visible energy losses typical of that interaction. $r - r_{\text{evt}}$ is the two-dimensional distance in the x, y-plane. For "Big Bird"-like events, the zenith and azimuth regions were further restricted for ν_μ-CC interactions to maximize the number of events with reconstructed values close to the observed ones, as the direction resolution is very good for tracks.

azimuth was chosen to cover a wide range of $\pm 110°$ to account for possible contributions from azimuthal regions affected by the ice anisotropy and due to the limited azimuthal resolution for single cascades. The zenith region was restricted more as the zenith resolution is better and spans $\pm 35°$. Both direction ranges are centered on the reconstructed directions. For "Big Bird"-like events generated as ν_μ-CC interactions, the zenith and azimuth regions were further restricted to a span of $\pm 17°$ and $\pm 40°$, respectively, to increase the statistics of events with reconstructed properties similar to "Big Bird", as the directional resolution for ν_μ-CC events is much better than for other event types. In case of the primary energy, the mapping depends on the neutrino spectrum and the interaction type, and is only well-correlated to the reconstructed deposited energy for ν_e-CC interactions, as only in this case the neutrino deposits its entire energy in the form of visible energy in the detector. All other interactions have some non-visible energy losses – final state neutrinos, intrinsically darker hadronic cascades, or muons leaving the detector – such that it is not a priori known what primary energy range will significantly contribute to the region around the reconstructed values. The primary neutrino energy was restricted such as to cover all energies that can contribute to the observed reconstructed energies, which had to be determined by trial and error. True values for the energy asymmetry and double cascade length are only defined for ν_τ-CC interactions. Those variables were left unconstrained during resimulation. Only "Double Double"-like events were resimulated as atmospheric muons; there are no atmospheric muons expected with energies of ~ 2 PeV passing the HESE selection.
After generation, the events had to pass the requirement for visible energy as shown in the second line of Table 7.1 to yield similar observed energies to the data events.

	"Big Bird"		
	generated	pass HESE & E_{vis}	double cascade topology
ν_e-CC	$1 \cdot 10^6$	$2.8 \cdot 10^5$	$7.2 \cdot 10^3$
ν_e-NC	$2 \cdot 10^6$	$2.1 \cdot 10^4$	$0.94 \cdot 10^3$
ν_e-GR	$0.4 \cdot 10^6$	$1.1 \cdot 10^4$	$0.47 \cdot 10^3$
ν_μ-CC	$4 \cdot 10^6$	$2.0 \cdot 10^5$	$0.68 \cdot 10^3$
ν_μ-NC	$2 \cdot 10^6$	$2.3 \cdot 10^4$	$1.0 \cdot 10^3$
ν_τ-CC	$2 \cdot 10^6$	$3.8 \cdot 10^5$	$206 \cdot 10^3$
ν_τ-NC	$2 \cdot 10^6$	$2.3 \cdot 10^4$	$1.0 \cdot 10^3$

Table 7.2: Resimulation statistics for "Big Bird". The number of total generated events (*left*), number of events passing the HESE selection and the visible energy requirement given in Table 7.1 (*center*), and number of events classified as double cascades (*right*) are shown for all classes of resimulated interactions.

	"Double Double"		
	generated	pass HESE & E_{vis}	double cascade topology
ν_e-CC	$10 \cdot 10^6$	$6.6 \cdot 10^6$	$76.6 \cdot 10^3$
ν_e-NC	$10 \cdot 10^6$	$0.5 \cdot 10^6$	$7.5 \cdot 10^3$
ν_e-GR	$2 \cdot 10^6$	$5.6 \cdot 10^3$	$0.12 \cdot 10^3$
ν_μ-CC	$40 \cdot 10^6$	$9.1 \cdot 10^6$	$72.6 \cdot 10^3$
ν_μ-NC	$10 \cdot 10^6$	$0.5 \cdot 10^6$	$7.7 \cdot 10^3$
ν_τ-CC	$10 \cdot 10^6$	$3.6 \cdot 10^6$	$183 \cdot 10^3$
ν_τ-NC	$10 \cdot 10^6$	$0.5 \cdot 10^6$	$7.7 \cdot 10^3$
atm. μ	$12 \cdot 10^9$	0	0

Table 7.3: Resimulation statistics for "Double Double". The number of total generated events (*left*), number of events passing the HESE selection and the visible energy requirement given in Table 7.1 (*center*), and number of events classified as double cascades (*right*) are shown for all classes of resimulated interactions..

The range for the visible energy is large compared to the statistics-only median resolutions on the visible energy of 8% for ν_e and ν_τ interactions and 11% for ν_μ interactions given in Section 4.2. Only the events passing the visible energy requirement were processed through the photon propagation and detector simulation stages. The events were then filtered and reconstructed in the same way as the initial MC events and the data were. The resimulation statistics are shown in Tables 7.2 and 7.3 for "Big Bird"-like and "Double Double"-like events, respectively. The first column shows the total number of generated events, the second column the number of events that pass the visible energy requirement and the HESE selection. None of the $12 \cdot 10^9$ generated muons pass the HESE selection. In total, $\sim 20 \cdot 10^6$ "Double-Double"-like events and $\sim 1 \cdot 10^6$ "Big-Bird"-like events from the dedicated resimulation pass the HESE selection. Those are

the statistics used in the a posteriori evaluation. For reference, the last column on the table shows the number of events getting classified as double cascades according to the scheme shown in Figure 5.2.
The distributions of the resimulated events are described in detail in Appendix B.

7.2 Estimating the Probability of a Tau-Neutrino Origin of the Double Cascades

The targeted MC aimed at the production of events with properties similar to these of the found double cascades. The resimulation can be used to properly quantify the probability that each double cascade was indeed induced by a ν_τ interaction. One possibility would be to evaluate the double cascades according to the original classification scheme, but that would neither take into account how close to the edge of the energy asymmetry range allowed for double cascades "Big Bird" is, nor the shape of the energy asymmetry distribution.
Let $\vec{\eta}_{\text{evt}}$ be the parameter space vector of each observed event containing its reconstructed properties: $\vec{\eta}_{\text{evt}} = (x, y, z, \cos(\theta_z), \phi, E_{\text{tot}}, A_E, L_{\text{dc}})_{\text{evt}}$. The posterior probability for each event to have originated from a ν_τ interaction is according to Bayes' theorem:

$$P(\nu_\tau \mid \vec{\eta}_{\text{evt}}) = \frac{P(\vec{\eta}_{\text{evt}} \mid \nu_\tau) P(\nu_\tau)}{P(\vec{\eta}_{\text{evt}})} \tag{7.1}$$

$$= \frac{P(\vec{\eta}_{\text{evt}} \mid \nu_\tau) P(\nu_\tau)}{P(\vec{\eta}_{\text{evt}} \mid \nu_\tau) P(\nu_\tau) + P(\vec{\eta}_{\text{evt}} \mid \cancel{\nu_\tau}) P(\cancel{\nu_\tau})}. \tag{7.2}$$

$P(\nu_\tau)$ and $P(\cancel{\nu_\tau})$ are given by:

$$P(\nu_\tau) = \frac{N_{\nu_\tau}^{\text{DC}}}{N_{\nu_\tau}^{\text{DC}} + N_{\cancel{\nu_\tau}}^{\text{DC}}},\ P(\cancel{\nu_\tau}) = \frac{N_{\cancel{\nu_\tau}}^{\text{DC}}}{N_{\nu_\tau}^{\text{DC}} + N_{\cancel{\nu_\tau}}^{\text{DC}}}, \tag{7.3}$$

where $N_{\nu_\tau}^{\text{DC}}$ and $N_{\cancel{\nu_\tau}}^{\text{DC}}$ are the total expected number of events classified as double cascades that are stemming from ν_τ and non-ν_τ interactions, respectively. These numbers can be taken from the original, full-detector MC simulation. Using the shorthand notation $P_{\nu_\tau}(\vec{\eta}_{\text{evt}})$ for $P(\vec{\eta}_{\text{evt}}|\nu_\tau)$ and $P_{\cancel{\nu_\tau}}(\vec{\eta}_{\text{evt}})$ for $P(\vec{\eta}_{\text{evt}}|\cancel{\nu_\tau})$, Equation 7.2 becomes:

$$P(\nu_\tau \mid \vec{\eta}_{\text{evt}}) = \frac{P_{\nu_\tau}(\vec{\eta}_{\text{evt}})\ N_{\nu_\tau}^{\text{DC}}}{P_{\nu_\tau}(\vec{\eta}_{\text{evt}})\ N_{\nu_\tau}^{\text{DC}}\ +\ P_{\cancel{\nu_\tau}}(\vec{\eta}_{\text{evt}})\ N_{\cancel{\nu_\tau}}^{\text{DC}}}. \tag{7.4}$$

P_{ν_τ} and $P_{\cancel{\nu_\tau}}$ are the PDFs for the ν_τ and non-ν_τ components evaluated at $\vec{\eta}_{\text{evt}}$. These are only crudely known when using the original MC sets and as mentioned before, can then

only be evaluated in two dimensions, as the full-detector MC by construction samples the entire possible observable parameter space, $\{\eta\}$. But $\{\eta\}$ around $\vec{\eta}_{\mathrm{evt}}$ is sampled much better for the two double cascades in the targeted resimulation, and contains a much more precise estimate of $P_{\nu_\tau}(\vec{\eta}_{\mathrm{evt}})$ and $P_{\not{\nu}_\tau}(\vec{\eta}_{\mathrm{evt}})$.

7.2.1 The RODEO Algorithm

The differential expected number of events at the point $\vec{\eta}_{\mathrm{evt}}$ is $N_{\nu_\tau} P_{\nu_\tau}(\vec{\eta}_{\mathrm{evt}})$ for ν_τ-induced events and $N_{\not{\nu}_\tau} P_{\not{\nu}_\tau}(\vec{\eta}_{\mathrm{evt}})$ for ν_e- and ν_μ-induced events. Note the removal of the DC-label, as the resimulated events are not required to pass the A_E requirements from the original classification. Instead, the distributions of A_E around the observed data will be evaluated, which for "Big Bird " happens to be very close to the boundary of the A_E region required for a double cascade classification.
The resimulation sets were generated on a parameter space that was restricted in multiple dimensions. Thus, an unbiased, computationally efficient method is needed to evaluate the resimulated events on the multi-dimensional parameter space. Owing to the high dimensionality of the problem, the distribution of the resimulated events may still be sparse in some dimensions. The method of choice must be robust to statistical fluctuations in the MC events' distributions.
The Regularization Of Derivative Expectation Operator (RODEO) was developed for problems similar to the one at hand. It is an algorithm computing the density estimate for sparse datasets in multiple dimensions. Specifically, it provides an unbiased and computationally efficient way to find the optimal bandwidth $\vec{h}$ in d dimensions for a $d-$dimensional set of n unweighted events. The optimal bandwidth is the kernel width giving the best, i.e. accurate and robust, density estimate for the given number and distribution of events. For each dimension, the optimal bandwidth h_d is chosen, accounting for that variable's relevance. The events are weighted by an appropriate kernel and the chosen bandwidth, giving the density estimate at the evaluated point $\vec{\eta}$. It can be seen as the unbinned, higher-dimensional version of evaluating the contributions of signal-like and background-like events, which has previously been carried out in a two-dimensional binned way.
To compute $N_{\nu_\tau} P_{\nu_\tau}(\vec{\eta}_{\mathrm{evt}})$ and $N_{\not{\nu}_\tau} P_{\not{\nu}_\tau}(\vec{\eta}_{\mathrm{evt}})$ for each event from the resimulation sets, the local version of the RODEO, yielding the optimal bandwidth at each of the evaluated points, is used with a gaussian kernel. The kernel density estimate (KDE) for a set of n

observables X in d dimensions at the point $\vec{\eta}$ is

$$f(\vec{\eta}, \vec{h}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \frac{1}{h_j} \mathcal{K}_j(\eta_j, X_{ij}, h_j) \tag{7.5}$$

where $\vec{h}$ is the n-dimensional bandwidth and $\mathcal{K}$ is a gaussian kernel,

$$\mathcal{K}_j = \frac{1}{\sqrt{2\pi h_j^2}} \exp\left(\frac{(\eta_j - X_{ij})^2}{2h_j^2}\right) \tag{7.6}$$

To use the RODEO, the events' properties need to be mapped to a $[0,1]^d$ space. Their distributions have to be reasonably well behaved in order to choose an appropriate kernel. The initial bandwidths $h_{j,0}$, $1 \leq j \leq d$ must be large such that random fluctuations in the properties of the MC events do not influence the density estimate. $\mathcal{A} = \{1, 2, ..., d\}$ is the initial set of dimensions. The derivatives Z_j of the kernel density estimate f_h with respect to each bandwidth h_j are computed and compared to their variance s_j:

$$Z_j = \frac{\partial f(\eta_j, h_j)}{\partial h_j} = \exp\left(\frac{(\eta_j - X_{ij})^2}{2h_j^2} \frac{(\eta_j - X_{ij})^2 - h_j^2}{h_j^4 \sqrt{2\pi}}\right), \tag{7.7}$$

$$s_j = \frac{1}{n} \mathrm{Var}(Z). \tag{7.8}$$

If $|Z| = |Z_{\not j} Z_j| > s_j c(n)$, where $c(n) \propto \sqrt{\log(n)}$ and $Z_{\not j} = \prod_{k \neq j} f_k(\eta_k, h_k)$ is the non-derivative part of the kernel density estimate, the bandwidth is reduced by a reduction factor β, with $0 < \beta < 1$. Here, $\beta = 0.95$ is chosen, as it allows a smooth reduction of the bandwidth without getting computationally overly expensive. If $|Z| < s_j c(n)$, the dimension j is removed from the set $\mathcal{A}$. This process is carried out iteratively, going around the dimensions, until $\mathcal{A}$ is an empty set.

The RODEO algorithm was developed for unweighted events, however, the MC events have weights assigned to them, which are not equal. Therefore, the RODEO algorithm needs to be extended to treat weights appropriately. This is accomplished by using the events' specific weights w_i, and by introducing the effective number of events $n_{\text{eff}} = (\sum_i w_i)^2 / \sum_i w_i^2$ and the effective weight $w_{\text{eff}} = \sum_i w_i^2 / \sum_i w_i$. This is the same treatment that is used in the SAY likelihood introduced in Section 4.2.1.

7.2.2 Applying the RODEO

Now, the dimensionality of the actual problem is revisited and the RODEO algorithm is applied to evaluate the probabilities for the found double cascades to stem from ν_τ interactions. There are the dimensions used in the original classification, the total reconstructed energy E_{tot} and the double-cascade length L_{dc}. Further, the energy asymmetry A_E should be elevated from a selection property to an analysis variable used in the determination of the posterior probability is explored. The resimulation was carried out in a restricted parameter space for E_ν, the three vertex dimensions *x, y, z* and the two direction dimensions θ, ϕ.
The L_{dc} observable has an exponentially falling distribution for ν_τ and ν_e, and is approximately flat for ν_μ, it further has no upper bound. The A_E observable has a range of $[-1, +1]$, but it has a flat distribution for ν_τ and ν_μ, and an approximately parabolic distribution for ν_e.
The range of *x, y, z* and θ, ϕ is limited by their resimulation volume. The bounds of the resimulation are defined as the bounds of the distribution, ignoring events with reconstructed properties beyond the resimulated volume. This is acceptable, as only events that have similar properties to the data events are of interest. Further, the resimulation ranges were chosen wide enough to allow for the inclusion of events in the evaluation that have different true values, but similar (mis-)reconstructed properties as the data events.
E_ν is not an observable, but the difficulties in mapping between the simulated quantity E_ν and the observable E_{tot} have already been discussed in Section 7.1. For E_{tot}, the ranges of the required visible deposited energy are chosen as the bounds of the volume. The evaluation ranges are thus equal to the resimulation ranges given in Table 7.1.
Then, the event properties are rescaled, such that the lower range corresponds to zero, and the upper range corresponds to one: $\min(X_{ij}) = 0, \max(X_{ij}) = 1$, where now $1 \leq j \leq d = 8$. L_τ and A_E were not restricted during the resimulation, L_{dc} is not bounded, and A_E has a flat distribution for ν_τ, a largely parabolic distribution for ν_e which further is spectrum-dependent and locally non-flat near the values measured for both "Big Bird" and "Double Double". Therefore, L_{dc} or A_E will not be used as a dimension in the RODEO but are removed from $\mathcal{A}$. Instead, a region of interest (ROI) is defined in those variables, the size of which is determined by the statistical error on the density estimates. The determination of the ROI is described in Section 7.2.3.
Note that as the events are weighted, the expressions for the kernel density estimate, Equation 7.5, as well as its derivative and variance, Equations 7.7 and 7.8, depend on the weights, and therefore on the assumed spectrum. Specifically, the astrophysical spectral

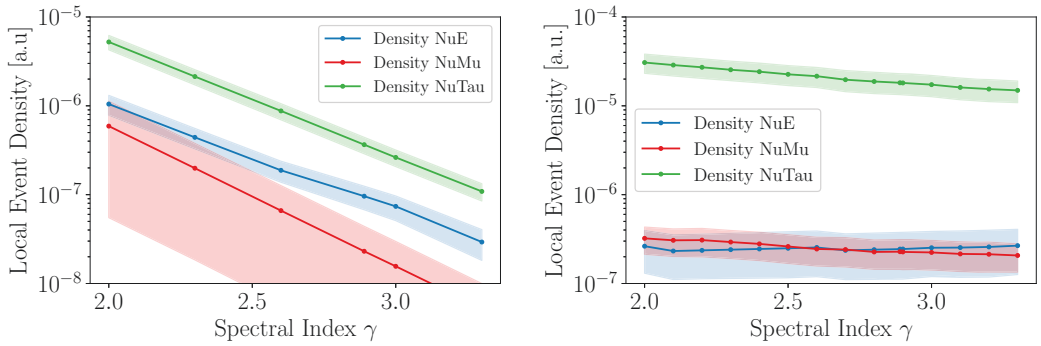

Figure 7.1: Local event density estimates evaluated using the RODEO algorithm for various values of the spectral index γ_{astro}. For "Big-Bird"-like events, the densities depend strongly on the spectral index of the astrophysical neutrino flux (*left*). For "Double-Double"-like events, the dependence on the spectral index is much weaker (*right*).

index γ and the flavor composition $\nu_e : \nu_\mu : \nu_\tau$ influences the relative contributions of the neutrino flavors. Figure 7.1 shows the local event density estimates as obtained with the RODEO algorithm for various values of the astrophysical spectral index γ_{astro}. "Big Bird" has a high reconstructed $E_{\text{tot}} = 1.8$ PeV and thus the expected number of observed events strongly depends on the spectral index. It can also be seen that for softer γ_{astro}, the ν_e contribution increases. As "Double Double " has a reconstructed $E_{\text{tot}} = 89$ TeV, which is very close to the anchor point of the astrophysical normalization at $E_\nu = 100$ TeV, the influence of the astrophysical spectral index γ_{astro} on the densities is not that large. However, as ν_τ-induced events have a larger contribution from larger E_ν for a given E_{tot} than ν_e-induced events, the relative contribution of ν_e-induced background increases with softening γ_{astro}.

7.2.3 Tauness

Equation 7.4 can be written as the ratio of the differential rates at $\vec{\eta}_{\text{evt}}$,

$$P(\nu_\tau \mid \vec{\eta}_{\text{evt}}) = \frac{N_{\nu_\tau} P_{\nu_\tau}(\vec{\eta}_{\text{evt}})}{N_{\nu_\tau} P_{\nu_\tau}(\vec{\eta}_{\text{evt}}) + N_{\cancel{\nu_\tau}} P_{\cancel{\nu_\tau}}(\vec{\eta}_{\text{evt}})} \equiv \tau. \tag{7.9}$$

In the last step the **tauness** τ is defined as the Bayesian posterior probability of an event to originate from a ν_τ interaction. It can be estimated as the fraction of events close to the reconstructed properties $\vec{\eta}_{\text{evt}}$ of the data events, which are expected to stem from ν_τ interactions. The tauness is a measure of how likely each of the double cascades is to have originated in a ν_τ vs. any other interaction. Note that the

tauness is always evaluated under prior assumptions for the physics parameters $\vec{\theta}$, e.g. for $\vec{\theta} = \hat{\vec{\theta}}$ maximizing $\mathcal{L} = \mathcal{L}^{\mathrm{C}}\mathcal{L}^{\mathrm{T}}\mathcal{L}^{\mathrm{DC}}$. Approximating $N_{\nu_\tau}P_{\nu_\tau}(\vec{\eta}_{\mathrm{evt}}) \approx \hat{f}_{\nu_\tau}(\vec{\eta}_{\mathrm{evt}}, \hat{h}_{\nu_\tau})$ and $N_{\not{\nu_\tau}}P_{\not{\nu_\tau}}(\vec{\eta}_{\mathrm{evt}}) = \sum_{\alpha=e,\mu} \hat{f}_{\nu_\alpha}(\vec{\eta}_{\mathrm{evt}}, \hat{h}_{\nu_\alpha})$, where $\hat{f}_{\nu_\alpha}(\vec{\eta}_{\mathrm{evt}}, \hat{h}_{\nu_\alpha})$ is the density of ν_α for the optimal bandwidth $\hat{h}_{\nu_\alpha}$ determined by the `RODEO` algorithm in the region of interest, the tauness becomes

$$\tau = \frac{\hat{f}_{\nu_\tau}(\vec{\eta}_{\mathrm{evt}}, \hat{h}_{\nu_\tau})}{\sum_{\alpha=e,\mu,\tau} \hat{f}_{\nu_\alpha}(\vec{\eta}_{\mathrm{evt}}, \hat{h}_{\nu_\alpha})}. \tag{7.10}$$

The region of interest for the L_{dc} and A_E variable is identified by varying the size of this two-dimensional region and evaluating the tauness and its uncertainty, while using the results of the fit with the flavor composition fixed to $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$ given in Equation 6.1 as the prior assumptions for the physics parameters. In Figure 7.2, the influence of the size of the ROI in $L_{\mathrm{dc}} - A_E$-parameter space is shown. The region is normalized to the final size of the chosen ROI. It corresponds to $\Delta A_E = 0.1$ and $\Delta \log_{10}(L_{\mathrm{dc}}) = 0.07$, or, $L_{\mathrm{dc}} \approx L_{\mathrm{dc,evt}}\,{}^{+2.9}_{-2.6}$ m. The top panels show the resulting tauness, and the bottom panels the relative uncertainty on the density which is influenced by the Poisson fluctuation of the contributing MC events. For "Big Bird", the tauness is locally robust against variations of the size of the ROI around the chosen size. At very small regions of interest, the MC fluctuations become large, and the density estimate becomes unstable. At very large regions of interest, the flavor-sensitive information gets washed out, both due to the exponentially increasing single-cascade background at lower L_{dc} values and the rapidly increasing single-cascade background towards larger A_E. Recall that for "Big Bird", $A_{E,\mathrm{evt}}$ is at the edge of the selection region for double cascades – the larger the contributing A_E is, the more single-cascade-like events contribute. At the chosen size of the ROI, the MC statistics are large enough for the contributing flavors ν_e and ν_τ to keep the influence of Poisson fluctuations on the tauness small. ν_μ-induced events do not contribute significantly to the background for any size of the ROI.

For "Double Double", the tauness is locally decreasing with increasing size of the ROI around the chosen size. However, a much smaller size cannot be chosen due to the limited number of contributing MC events, especially from ν_μ-interactions. In case of "Double Double", ν_e and ν_μ events contribute approximately equally to the total background. With increasing size of the ROI, the number of single-cascade-like background increases. This is due to the rapid increase of single-cascade background, both towards lower L_{dc} values, as well as towards lower A_E values. A size of the ROI of twice the chosen size corresponds to $\Delta A_E = 0.2$, the evaluated region includes the physical bound of $A_E = -1$ and thus the region of $-1.0 \leq A_E < -0.98$. Due to an accumulation of true single cascades at the lower bound of A_E, this region was not part of the initial selection

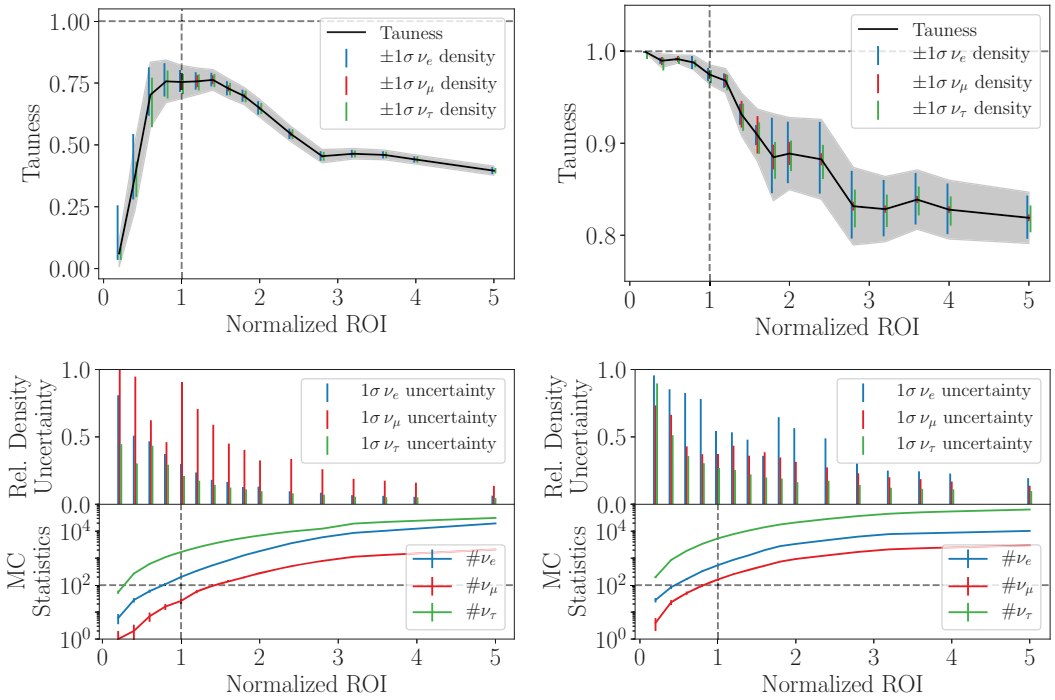

Figure 7.2: Tauness as a function of the size of the ROI in the $L_{\mathrm{dc}} - A_E$ parameter space for "Big Bird" (*left*) and "Double Double" (*right*). The tauness with uncertainty is shown together with the contributions from each of the flavors' densities (*top*). The relative density uncertainty and the number of Monte Carlo events used to evaluate the density are shown (*bottom*). The ON cut factor is normalized to the chosen ROI

for double cascades. As the choice of the size of the ROI is driven by the MC fluctuations in the ROI, the chosen size leads to a conservative estimate of the tauness of "Double Double" .

As shown in Figure 7.1, the estimated local densities depend onthe spectral index γ_{astro}. Therefore, also the tauness in Equation 7.10 depends on the spectrum, which is shown in Figure 7.3. For "Big-Bird", the influence of the spectral index on the tauness is within the statistical uncertainties. For "Double-Double", the tauness depends much stronger on the spectral index.

The tauness for each of the events to stem from a ν_τ-CC interaction is computed for the spectral parameters obtained from the fit with the flavor composition fixed to $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$ given in Equation 6.1. The tauness for "Big Bird" is $\tau^{\mathrm{BB}}_{\text{best fit}} \approx 75\%$, the tauness for "Double Double" is $\tau^{\mathrm{DD}}_{\text{best fit}} \gtrsim 97\%$. That means, ~ 75 ($\gtrsim 97$) out of 100 observed events with Big Bird (Double Double) reconstructed properties are expected

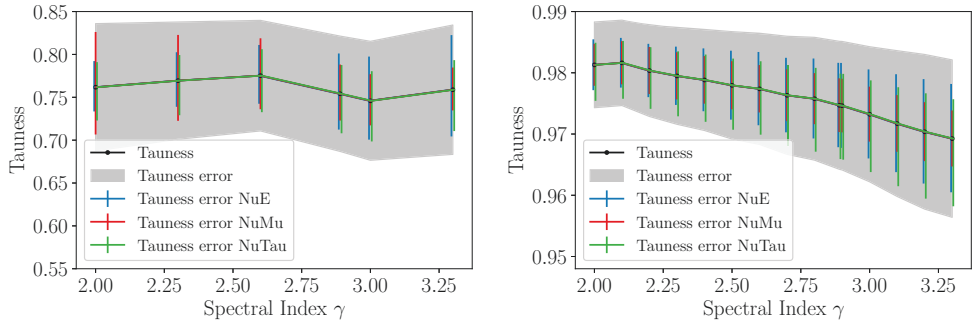

Figure 7.3: Tauness evaluated using densities obtained with the `RODEO` algorithm for various values of the spectral index γ_{astro} for "Big-Bird" (*left*) and "Double-Double"(*right*).

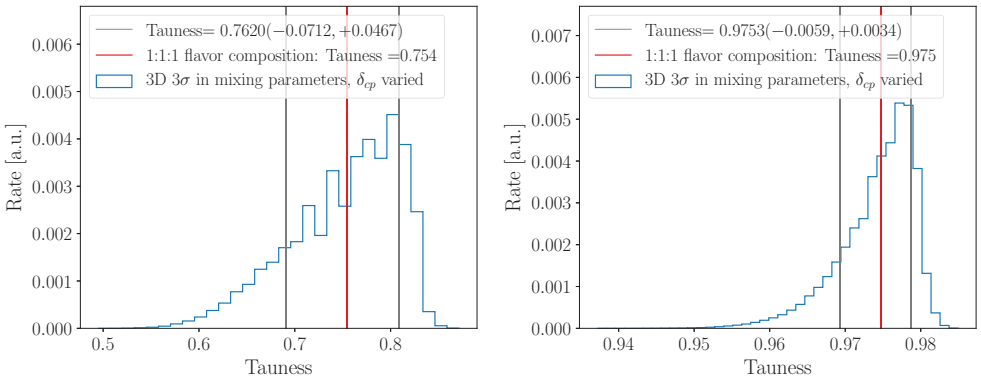

Figure 7.4: Distribution of tauness values for "Big Bird" (*left*) and "Double Double" (*right*) when varying the source flavor composition and the mixing parameters in the `NuFit4.1` allowed 3σ confidence interval.

to stem from ν_τ-CC interactions. For Double Double, the statistics of the generated MC are not sufficient to evaluate the tauness to a higher precision.

One can extend the tauness to take the uncertainties in the neutrino mixing matrix into account, and sample from the resulting PDF in the flavor triangle according to the uncertainties in the production mechanism, and the mixing parameters. Figure 7.4 shows the distribution of the tauness when uniformly sampling the entire source flavor composition $\nu_e : \nu_\mu : \nu_\tau$ and varying the mixing parameters in the 3σ interval allowed by `NuFit4.1`, in the same way as presented in Section 2.3.3. As can be seen from Figure 7.4, the tauness for "Double Double" is $(97.5^{+0.3}_{-0.6})\%$when keeping the astrophysical spectral index fixed at $\gamma = 2.88$ and varying the source flavor composition over the entire parameter space and the mixing parameters in their 3σ allowed range. For "Big Bird", the tauness is $(76.2^{+4.7}_{-7.1})\%$ for the same variation of parameters.

7.3 Updated Flavor Composition Measurement

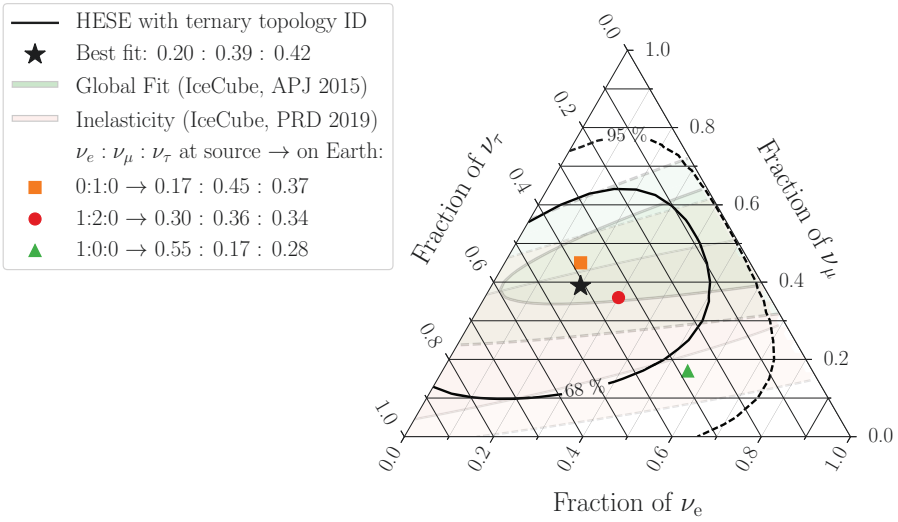

Figure 7.5: Results of the flavor composition measurement using the updated likelihood for double cascades containing the probability densities obtained from the RODEO algorithm. Contours show the 1σ and 2σ confidence intervals assuming Wilks' theorem holds. The shaded regions show previously published results [32, 154] without direct sensitivity to the tau-neutrino component. Flavor compositions expected from popular astrophysical neutrino production mechanisms are marked.

The densities obtained with the RODEO algorithm are used to perform the flavor composition measurement from Section 6.1 again, with the additional information included. The previously used total likelihood for the three topologies single cascade, track, and double cascade is $\mathcal{L} = \mathcal{L}^{\mathrm{SC}}\mathcal{L}^{\mathrm{T}}\mathcal{L}^{\mathrm{DC}}$ where $\mathcal{L}^{\mathrm{SC,T}}$ are the HESE binned-likelihood over the single cascade and track events respectively, as given in Equation 5.3 and used before. The likelihood uses the two-dimensional MC PDFs based on total deposited energy E_{tot} and cosine of the zenith angle $\cos(\theta_z)$ for single cascades and tracks. $\mathcal{L}^{\mathrm{DC}}$ is now replaced by the extended, unbinned likelihood for the double cascades, which is

$$\mathcal{L}^{\mathrm{DC}}_{\mathrm{Rodeo}} = e^{-\sum_c N_c} \prod_{\mathrm{evt}} \left(\sum_c N_c P_c(\vec{\eta}_{\mathrm{evt}}) \right), \tag{7.11}$$

where c are the components used in the fit, $c = \nu_{\mathrm{astro},\alpha}, \nu_{\mathrm{conv},\alpha}, \nu_{\mathrm{prompt},\alpha}, \mu_{\mathrm{atm}}$ for the flavors $\alpha = e, \mu, \tau$. $N_c P_c(\vec{\eta}_{\mathrm{evt}})$ is computed using the RODEO algorithm, with the dependence on γ_{astro} parametrized.

The results of the flavor composition measurement using the extended unbinned likeli-

	Single power law		
	$\nu_e : \nu_\mu : \nu_\tau$ fixed	$\nu_e : \nu_\mu : \nu_\tau$ free	$\nu_e : \nu_\mu : \nu_\tau$ free
likelihood $\mathcal{L}$	$\mathcal{L}^{\mathrm{SC}}\mathcal{L}^{\mathrm{T}}\mathcal{L}^{\mathrm{DC}}$	$\mathcal{L}^{\mathrm{SC}}\mathcal{L}^{\mathrm{T}}\mathcal{L}^{\mathrm{DC}}$	$\mathcal{L}^{\mathrm{SC}}\mathcal{L}^{\mathrm{T}}\mathcal{L}^{\mathrm{DC}}_{\mathrm{Rodeo}}$
Φ_{astro}	6.37 $[-1.6, +1.5]$	6.98 $[-2.1, +2.4]$	7.35 $[-2.1, +2.4]$
γ_{astro}	2.87 $[-0.19, +0.20]$	2.89 $[-0.20, +0.21]$	2.87 $[-0.20, +0.21]$
$\vec{f}_\alpha$	1/3 : 1/3 : 1/3	0.29:0.43:0.28	0.20:0.39:0.42
Φ_{conv}	1.01 $[-0.34, +0.35]$	0.97 ± 0.36	0.97 ± 0.36
Φ_{prompt}	0.0 $[-0.0, +5.3]$	0.0 $[-0.0, +5.2]$	0.0 $[-0.0, +6.1]$
Φ_μ	1.19 $[-0.45, +0.46]$	1.16 $[-0.45, +0.46]$	1.16 $[-0.45, +0.46]$
$\Delta\gamma_{\mathrm{CR}}$	-0.05 $[-0.13, +0.05]$	-0.05 $[-0.13, +0.05]$	-0.05 $[-0.13, +0.05]$
$R_{K\pi}$	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.01
$2\nu/(\nu+\bar{\nu})_{\mathrm{atm}}$	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.01
ϵ_{DOM}	0.95 $[-0.07, +0.09]$	0.95 $[-0.07, +0.10]$	0.94 $[-0.07, +0.06]$
$\epsilon_{\text{head-on}}$	-0.06 $[-0.48, +0.51]$	-0.05 $[-0.48, +0.51]$	-0.05 $[-0.48, +0.52]$
a_s	1.0 ± 0.2	1.0 ± 0.2	1.0 ± 0.2

Table 7.4: Comparison of the Results for the flavor composition fits assuming a single-power-law. The left column shows the results obtained when assuming a 1 : 1 : 1 neutrino flavor composition. The center column shows the results when fitting for the neutrino flavor composition using two-dimensional PDFs for all topologies. The right column shows the results when using the targeted MC simulation and the unbinned likelihood for double cascades. The fit parameters are explained in Table 5.3.

hood for double cascades is shown in Figure 7.5. The best-fit point is

$$\frac{\mathrm{d}\Phi_\nu}{\mathrm{d}E} = 7.34 \cdot \left(\frac{E}{100\ \mathrm{TeV}}\right)^{-2.87} \cdot 10^{-18} \cdot\ \mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1} \tag{7.12}$$

with a flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 0.20 : 0.39 : 0.42$. As can be seen, the best-fit point moved along the $\nu_e - \nu_\tau$ axis wrt. the measurement solely relying on two-dimensional information and presented in Section 6.1. This is due to the fact that in the original two-dimensional evaluation, both events had quite similar tauness values of $\tau_{\mathrm{BB}} \approx 0.5$ and $\tau_{\mathrm{DD}} \approx 0.6$, and could be described by a ν_τ component, but also a ν_e component fairly well. With the multidimensional evaluation, the ν_e and ν_μ hypotheses are compatible for "Double Double" only at $\sim 1.5\%$ each. "Double Double"'s high probability of being a ν_τ drives the flavor composition towards a higher ν_τ fraction, and leads to a ν_τ fraction of $f_{\nu_\tau} = 0$ to be disfavored at $> 2\sigma$. Comparing this result with previously published results of the flavor composition clearly shows the advantages of the ternary topology classification. Not only is the best-fit point non-zero in all flavor components for the first time, but the degeneracy between the ν_e and ν_τ fraction is

broken. Also the influence of the small number of tracks in the HESE sample can be seen, which leads to an increased uncertainty on the ν_μ fraction as compared to [32] and [154]. This measurement is compatible with the previously published results. The green shaded contours in Figure 7.5 show the results of a global-fit to multiple samples [32], where the flavor composition measurement was performed using only the distinction between cascades and tracks. This leads to the aforementioned degeneracy along the $\nu_e - \nu_\tau$-axis. The larger number of tracks present in the combined sample, however, constrains the ν_μ-fraction better. The red shaded contours show the results of an inelasticity measurement [154] performed on a sample containing only starting tracks and extending to lower energies. The inelasticity distribution is expected to peak at different values for ν_μ-CC events than for ν_τ-CC events where the tau decays into a muon, due to the invisible energy carried away by the second ν_τ produced in the tau decay. The inelasticity measurement thus offers some sensitivity to the relative ν_μ-to-ν_τ ratio, even at lower energies; but using only tracks, it leaves the ν_e-fraction unconstrained. Even though both these measurements have very different best-fit flavor compositions of $\nu_e : \nu_\mu : \nu_\tau = 0.5 : 0.5 : 0.0$ for the global-fit [32] and $\nu_e : \nu_\mu : \nu_\tau = 0.0 : 0.2 : 0.8$ for the inelasticity measurement [154], they are compatible with each other due to their limited sensitivity.

A comparison of the single-power-law fits performed on the HESE-7 sample is shown in Table 7.4. The first column shows the results of the fit with a fixed flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 1$, the second column shows the results of the flavor fit using two-dimensional PDF templates for each flavor, the results in the third column are from the flavor fit using the extended multidimensional information for the found double cascades. The fit spectral index is very similar for the three measurements, which is expected as the spectral index is assumed the same for all flavors. The total astrophysical normalization shows a larger variation, which is still well within the measurement uncertainty. This variation is expected, as fitting for the total astrophysical normalization with the two additional parameters describing the relative per-flavor contributions to the total normalization, is equivalent to fitting for three independent per-flavor normalizations. The total normalization is the parameter most affected by allowing the flavor composition to deviate from $1 : 1 : 1$. As for the parameters describing systematic uncertainties, the treatment and sensitivity is not changed by fitting the flavor composition. The only exception is the ice anisotropy, which affects the double cascade length observable, and thus the flavor composition measurement. However, it has been shown in Section 5.4 that deviations of 20% on the total scale of the anisotropy strength, given by the different ice models *SpiceLea*, for which the anisotropy is corrected for in reconstruction, and

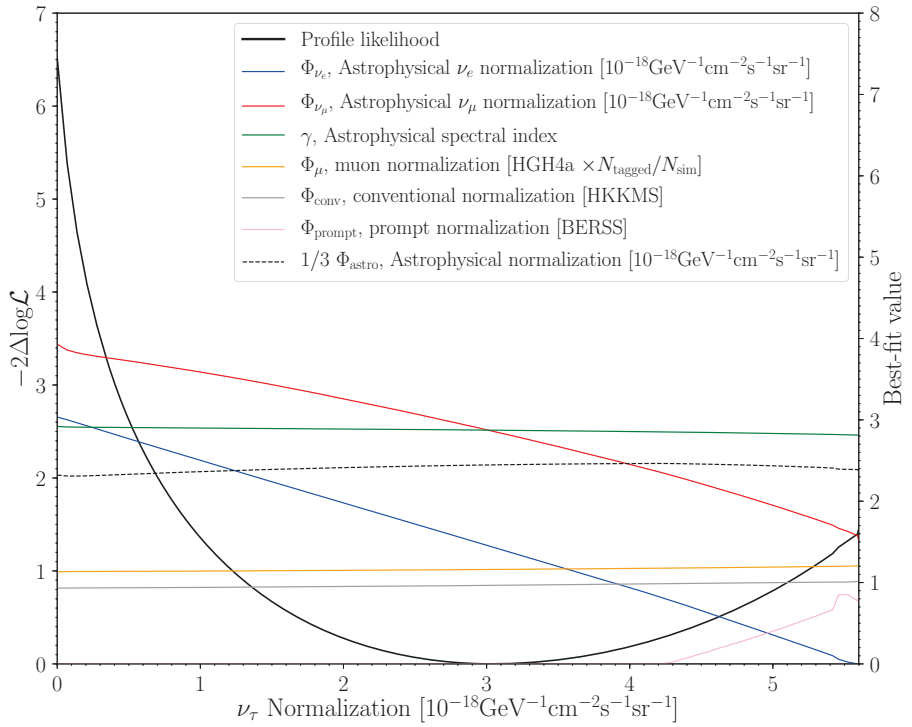

Figure 7.6: Profile likelihood scan of the astrophysical ν_τ normalization using the updated likelihood for double cascades containing the probability densities obtained from the RODEO algorithm. Assuming Wilk's theorem holds, a zero ν_τ flux is disfavored at 2.8σ.

Spice3.2, which the MC events are simulated with, has no influence on the misclassifaction fraction. Further, none of the double cascades has a reconstructed direction that would be affected by scaling the ice anisotropy. This further limits the sensitivity to fit the anisotropy from data, and the influence of this particular ice optical parameter on the final result.

Using the likelihood in Equation 7.11, a one-dimensional scan of the astrophysical ν_τ flux is performed, and shown in Figure 7.6. All other components of the fit are left free, and assume their conditional best-fit values. It can be seen that while the relative ν_e and ν_μ normalizations do change when varying the ν_τ normalization, the remaining fit parameters stay approximately constant. Only when the ν_τ component dominates the astrophysical flux, a non-vanishing prompt component is needed to account for the single cascades observed at high energies. The test statistic $-2\Delta\log\mathcal{L}$ compares the best-fit value at $\Phi_{\nu_\tau} = 3.0 \cdot 10^{-18}\,\text{GeV}^{-1}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}$ to each scanned point. In particular, the

test statistic

$$\mathrm{TS} = -2\Delta \log \mathcal{L} = -2 \log \left(\frac{\mathcal{L}(\vec{n}_{\mathrm{evt}} \mid \vec{\mu}(\Phi_{\nu_\tau} = 0, \hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\xi}}}))}{\mathcal{L}(\vec{n}_{\mathrm{evt}} \mid \vec{\mu}(\hat{\Phi}_{\nu_\tau}, \hat{\vec{\theta}}, \hat{\vec{\xi}}))} \right), \tag{7.13}$$

comparing the likelihood of a fit with a ν_τ flux fixed at $\Phi_{\nu_\tau} = 0$ to the best fit Φ_{ν_τ}, gives the significance at which a vanishing astrophysical tau neutrino flux can be disfavored. Wilks' theorem [151] holds when the parameters of interest are not bounded, which is not true here. Equation 7.13 is evaluated at a physical boundary, given that the tau neutrino flux cannot be negative. It has been shown [152] that in this case the test statistic is expected to follow a half-$\chi^2_{k=1}$ distribution, $0.5\delta(\lambda) + 0.5\chi^2_{k=1}(\lambda)$, with $k = 1$ degree of freedom. As can be read off from the y-intercept of the profile likelihood in Figure 7.6, the observed test statistic is $\lambda = TS = 6.5$, which translates to a significance of 2.78σ, or a p-value of 0.5%.

The 1σ confidence intervals are found by requiring

$$\mathrm{TS} = -2\Delta \log \mathcal{L} = -2 \log \left(\frac{\mathcal{L}(\vec{n}_{\mathrm{evt}} \mid \vec{\mu}(\Phi_{\nu_\tau}^{\mathrm{test}}, \hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\xi}}}))}{\mathcal{L}(\vec{n}_{\mathrm{evt}} \mid \vec{\mu}(\hat{\Phi}_{\nu_\tau}, \hat{\vec{\theta}}, \hat{\vec{\xi}}))} \right) \stackrel{!}{=} 1, \tag{7.14}$$

where the full $\chi^2_{k=1}(\lambda)$ distribution is used, as the evaluated point is not at the boundary. The value of the test statistic $\lambda = TS = 1$ is obtained for $\Phi_{\nu_\tau} = [1.2, 5.2] \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$, and the measurement of the astrophysical tau-neutrino flux normalization is measured to

$$\Phi_{\nu_\tau} = 3.0^{+2.2}_{-1.8} \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}. \tag{7.15}$$

Together with the spectral index $\gamma_{\nu_\tau} = \gamma_{\mathrm{astro}} = 2.87^{+0.21}_{-0.20}$, this constitutes the first non-zero measurement of the astrophysical tau-neutrino flux.

The validity of Wilk's theorem is assessed by performing pseudo MC experiments. Each pseudo MC experiment constitutes one realization of the experiment, based on the simulated data and the tested model. Events are drawn from the MC according to the injected model. Therefore, pseudo MC tests can only be performed using the original, untargeted MC sets containing all possible realizations of the tested model. The targeted MC has been generated for the one specific realization observed in nature, namely for the classification of two events as double cascades based on their reconstructed properties. The performed pseudo MC experiments for the untargeted MC are found to follow the expected distributions, validating the usage of Wilk's theorem. There is no reason

Wilk's theorem would hold over the entire range of possible realizations, but be violated in the specific observed realization. The tests are discussed in detail in Appendix C.

7.4 Source Flavor Composition

None of the source flavor scenarios discussed in Section 2.2.1 can be excluded with high significance even using the multidimensional likelihood-ratio test introduced in the previous section. The presented measurement is consistent with the expectation of astrophysical neutrinos being produced in pion-decays with a source flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 2 : 0$, and arriving on Earth with a flavor composition of $\nu_e : \nu_\mu : \nu_\tau \approx 1 : 1 : 1$. It is also consistent with a muon-decay production scenario with a source flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 0 : 1 : 0$, and a flavor composition on Earth of $\nu_e : \nu_\mu : \nu_\tau = 0.17 : 0.45 : 0.37$. The best-fit point is closest to the expectation if the muon-decay production scenario is assumed. But this result is consistent with both the pion- and muon-decay scenarios within 1σ. The neutron-decay scenario with a source flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 0 : 0$ and a flavor composition on Earth of $\nu_e : \nu_\mu : \nu_\tau = 0.55 : 0.17 : 0.28$ is mildly disfavored, but consistent at 2σ. Also, the charm production scenario with a source flavor composition of $\nu_e : \nu_\mu : \nu_\tau = 1 : 1 : 0$ and a flavor composition on Earth of $\nu_e : \nu_\mu : \nu_\tau = 0.36 : 0.31 : 0.33$ is consistent at 1σ.
However, exotic scenarios, leading to e.g. less neutrino oscillations during propagation, can be constrained. One such scenario involves high-dimensional terms from a higher-energy scale theory, leading to effective operators [**?**], which has been tested using the HESE-7 dataset [145]. Also interactions with Dark Matter could alter neutrino propagation and change the resulting flavor composition [155]. With this measurement, the total suppression of neutrino oscillations in either of these scenarios can be excluded at 2σ for neutrino production from pion-decay, and at $> 3\sigma$ for neutrino production from neutron or muon decay.

Chapter 8

Summary and Outlook

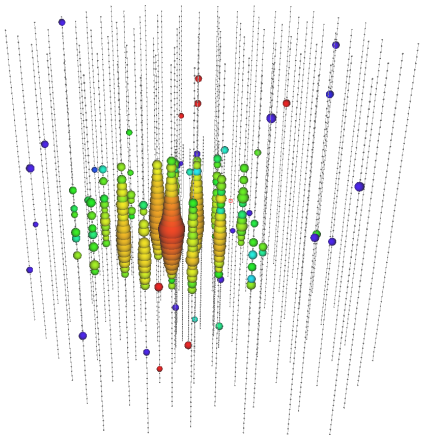

Figure 8.1: The high-confidence identified astrophysical tau neutrino, "Double Double", in IceCube's event viewer.

In this work, an update to the tau-neutrino search and flavor composition measurement using the High-Energy Starting Events sample has been performed. It includes the addition of 1.5 years of livetime, for a total of 7.5 years of livetime, an updated treatment of systematic uncertainties, and a re-processing of all previously taken data incorporating an improved detector calibration. A new calculation of the atmospheric self-veto, crucial to estimate the atmospheric contamination of the HESE sample, has also been incorporated. Previously a stand-alone analysis performed on top of the HESE selection [37], the reconstruction and subsequent topology classification have now become an inherent part of the HESE sample, superseding the previously-used reconstruction and by-eye classification into tracks and cascades [132]. This analysis adds the third topology of double

cascades to HESE, and an algorithmic classification, which is compared to Monte-Carlo simulation, so that mis-classification probabilities are quantified. The algorithmic classification employs event properties from the reconstruction of all possible topologies, a single cascade, a double cascade, and a track. It aims at a high purity in each topological sub-sample. Two-dimensional Monte Carlo PDF templates are employed for each topology and each flavor to fit the overall flavor composition of astrophysical neutrinos. In the double cascade topology, events stemming from ν_τ-CC interactions show a correlation between their reconstructed total deposited electromagnetic-equivalent energy and their reconstructed double-cascade length, which is not the case for misclassified NC, ν_e-CC- and ν_μ-CC-induced events. As a true double cascade topology is unique to tau-neutrino interactions, the ν_e-ν_τ degeneracy can be broken at high energies ($E \gtrsim 100$ TeV), where the double cascade topology becomes resolvable in IceCube. In combination with the single-cascade and track topologies, direct sensitivity to each neutrino flavor is obtained. For the fit, a new likelihood is used which accounts for the statistical uncertainties from the finite number of available Monte Carlo events.

By fully reconstructing the event topologies, two events have been classified as double cascades for the first time. They are IceCube's first prime candidates for astrophysical tau neutrinos. Both events have been previously known, but had thus far been classified as single cascades. The first event is the highest-energy HESE event, nicknamed "Big Bird". It has a reconstructed total electromagnetic-equivalent energy of 1.8 PeV and a reconstructed double-cascade length of 16 m. Within this region of the two-dimensional parameter space, the expectation for both signal and background is low. Furthermore, the energy asymmetry of 0.29 lies just within the range required for double cascades of ≤ 0.30. When the brightest, non-saturated DOMs are included in the reconstruction, "Big Bird" is reclassified as a single cascade. This might suggest that "Big Bird" might be a misclassified true single cascade. The second event has a reconstructed total electromagnetic-equivalent energy of 89 TeV and a reconstructed double cascade length of 17 m. This places it within a parameter space region which is occupied by both signal (ν_τ-induced) and background (ν_e- or ν_μ-induced) events. It has a larger double-cascade length than most misclassified single-cascade background events. Its energy asymmetry is -0.80, which lies in a signal-dominated region. Further, upon visual inspection, the observed light arrival pattern at the DOMs shows a good agreement with the light arrival pattern expected from a double cascade, and poor agreement with that expected from a single cascade.

To firmly conclude how likely a ν_τ interaction is for each of the found double cascades, an a posteriori analysis of these two events has been performed. Many MC events have

been generated with reconstructed observables similar to the data events. The MC events have been simulated in a restricted parameter space to provide high statistics with computational efficiency. All neutrino flavors and interactions have been simulated, and for "Double Double", single atmospheric muons have been also simulated. To obtain multidimensional PDFs for the resimulated parameter space, a kernel density estimation method for high-dimensional spaces has been employed, modified to account for the intrinsic weights of the MC events and limited MC statistics. The specific algorithm employed provides a computationally efficient method for the unbiased evaluation of sparse data in multiple dimensions.

The a posteriori analysis revealed that "Big Bird" is consistent with being a misclassified single cascade at the 25% level, and could have been induced by either a ν_τ, or a ν_e interaction. The second double cascade is consistent with indeed being induced by a ν_τ-CC interaction, with all background interactions having a combined probability of 3%. This event has been nicknamed "Double Double". "Double Double" is thus the first positively-identified astrophysical tau neutrino, which marks the finding of the long sought-after, last Standard Model cosmic messenger. An event view, using IceCube's visualization software, is shown in Figure 8.1. The two cascades are not identifiable by eye, highlighting the need for the algorithmic topology classification employed in this work. Tau neutrinos have a very high probability to be astrophysical, making them an interesting class of events for source searches. For the point source search using the HESE sample, each event was fit thousands of times, varying the directional hypotheses, but only using a binary topology classification. "Double Double" is not sufficiently well localized to allow for a counterpart search in the electromagnetic spectrum. In future, a double-cascade hypothesis could be incorporated into the sky scans, making use of the lever arm provided by the cascade separation and allowing for more precise reconstructions.

While the a posteriori analysis was ongoing, two analyses using the "double pulse" method to search for tau neutrinos have been performed. Both identify "Double Double" as a candidate tau-neutrino event [156, 157], further strengthening its case. While in the analysis presented here, the main remaining background for "Double Double" are single cascades, the main background for "double pulse" searches are ν_μ-induced tracks, making these two search techniques complementary.

The likelihood used for the flavor composition measurement was replaced for the double cascade topology, while keeping the binned likelihood using two-dimensional PDFs for tracks and cascades. The new, unbinned double cascade likelihood allows for the incorporation of a multi-dimensional PDF as evaluated by a kernel density estimator.

Using this new composite likelihood, a flavor composition measurement assuming a single power law for astrophysical neutrinos was performed. The best-fit flavor composition of astrophysical neutrinos is $\nu_e : \nu_\mu : \nu_\tau = 0.20 : 0.39 : 0.42$, which for the first time is non-zero in all flavor components. Likewise for the first time, a zero tau contribution can be disfavored, in this case at 2.8σ significance (p-value 0.5%). Further, the first positive measurement of the astrophysical tau-neutrino flux is provided, which has a normalization of $\Phi_{\nu_\tau} = 3.0^{+2.2}_{-1.8} \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ at 100 TeV and a spectral index of $\gamma_{\mathrm{astro}} = 2.87^{+0.21}_{-0.20}$. In this work, the first high-confidence identification of astrophysical tau-neutrino interactions and the first non-zero measurement of the astrophysical tau-neutrino flux have been achieved, but the uncertainties on the flux and the flavor composition are nonetheless still substantial. Our measured flavor composition is consistent with all previous measurements within 3σ and with neutrino production from pion or muon decay within 1σ and for neutrino production from neutron decay within 2σ, if standard neutrino oscillations are assumed. Constraints can be derived for new physics scenarios which completely inhibit neutrino oscillations during propagation. These are excluded at $> 3\sigma$ for neutrino production from neutron or muon decay, and disfavored at 2σ for pion production. This work, while of limited precision, is a first step to extend IceCube's neutrino oscillation studies to unprecedented energies and baselines.

The precision of the flavor composition measurement can be improved in the near future. The measurement presented here suffers from the low number of HESE events with reconstructed energies above 60 TeV. It has previously been shown that a combination of high-energy neutrino samples can greatly reduce the statistical errors on the flavor composition measurement, which was only performed using the two topologies of cascades and tracks [32]. An update of this so-called "global fit" is in preparation. The inclusion of through-going track-like, mainly ν_μ-induced, events would greatly decrease the errors on the ν_μ normalization. The addition of mainly ν_e- and ν_τ-induced cascades, selected by topology instead of an incoming track-veto, provides a larger sample of potential double cascades. Initial studies have shown that the combination of HESE with contained cascades would increase the number of identifiable tau neutrinos by $\approx 20 - 45\%$, depending on the underlying astrophysical spectral index [158]. Further gains can be expected from improvements in the events' classification. A per-event test statistic has been developed which is based on the likelihoods of the double cascade reconstruction. It can be used to better separate single cascades from double cascades, but its incorporation into the topology classification chain is beyond the scope of this thesis. A combination with the "double pulse" identifier can potentially be used to further boost the significance of candidate events. In this work, we have shown that kernel density estimators can

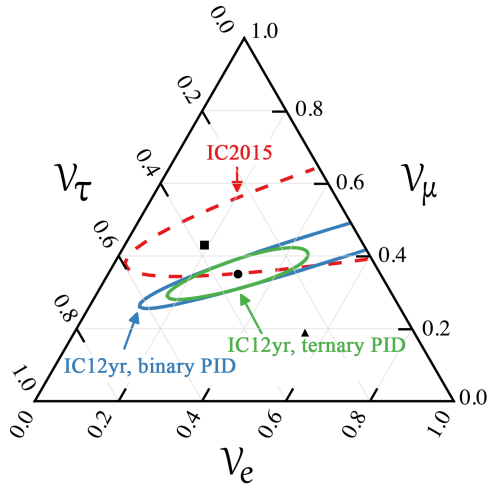

Figure 8.2: Future improvements on the flavor composition measurement. The red contour shows the previously published result using a binary classification and a combination of multiple high-energy neutrino samples [32]. The green contour shows the projected sensitivity after 12 years of data taking, using a combination of multiple high-energy neutrino samples and the ternary classification used in this work. All contours show 1σ confidence levels, with the projections further assuming no systematic uncertainties and perfect knowledge of atmospheric contributions.

be employed to evaluate events on a multidimensional parameter space, and for sparse and weighted MC events. This can serve a future flavor analysis by reducing selection criteria for classification, replacing them with multidimensional PDFs.

As far as systematic uncertainties are concerned, the ice optical properties have the largest influence on event reconstruction and classification. For the IceCube-Upgrade [159], new instrumentation, including newly-developed calibration devices, will be deployed into the Antarctic ice in the central region of IceCube in the next few years with horizontal spacings between modules between 20 m and 70 m. Using the LED-flashers on the new modules, experimental double-flasher events an be constructed with double-flasher lengths relevant to tau-neutrino searches; and the double-cascade-length resolution verified experimentally at multiple lengths. The gain in knowledge of the ice optical properties is expected to be substantial and will be of utmost importance for neutrino oscillation studies from the GeV to the PeV range. Figure 8.2 shows the estimated sensitivity of a future flavor composition measurement, assuming full control over systematic uncertainties. While this projection is unachievable in reality, both the global fit and the improved knowledge of ice optical properties will lead to a significant

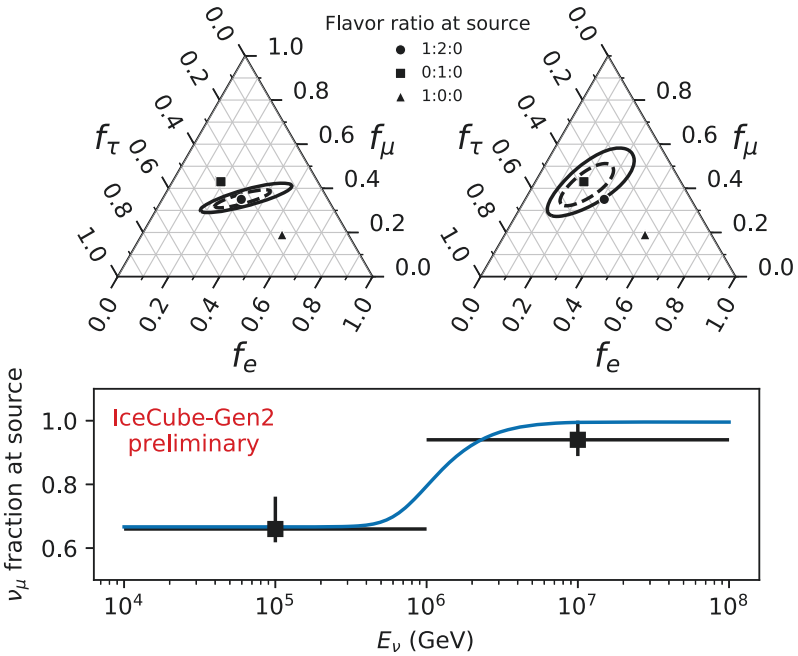

Figure 8.3: Future improvements on the flavor composition measurement with IceCube-Gen2, taken from [160]. The top panels show the flavor composition measured on Earth with 1σ and 2σ contours, the bottom panels the fraction of ν_μ produced at the sources. A transition from a pion production to a muon production is assumed at neutrino energies of 1 PeV.

improvement of the flavor composition measurement.

In the further future, the planned IceCube-Gen2 facility [160] will provide an instrumented volume that is several times larger than IceCube. Figure 8.3 shows the projected sensitivity of the flavor composition measurement, incorporating a study on the influence of the wider string spacing on tau-neutrino identification efficiencies [161]. A change from one neutrino production scenario to another at a certain energy scale may be resolvable. In the projection shown, a sharp transition from a pion-decay to a muon-decay neutrino production at neutrino energies of 1 PeV is assumed. Gen2 will use optical sensors with better angular acceptance, more light sensitive area and a segmented design. The new optical sensors will facilitate angular reconstruction of cascades, and should further increase the efficiency of tau-neutrino identification. The improved angular resolution of double cascades will not only benefit the tau-neutrino identification efficiency; as tau neutrinos have a negligible atmospheric contribution, they offer an almost background-free signal for follow-up observations in the electromagnetic spectrum and will most likely be of great interest for future neutrino source searches.

Appendix A

Reconstruction Stability

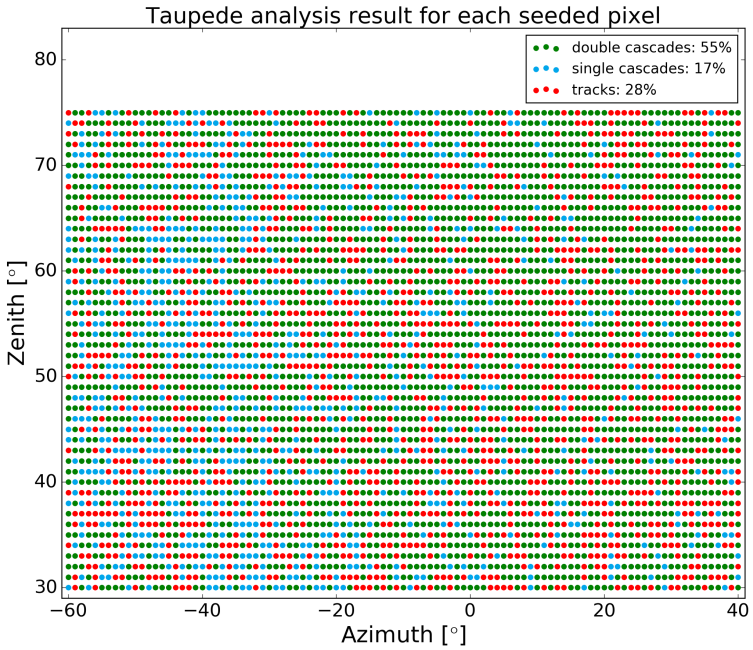

Figure A.1: Scan of the influence of the seed direction on the final classification of "Double Double". For directions spanning 45° in zenith and 100° in azimuth, the direction of the seed for `taupede` is fixed in 1° increments, and the event classified according to the procedure described in 5.2, with the resulting classification shown.

As shown in 5.3 this analysis has a limited tau neutrino identification efficiency of $< 10\%$ and almost 50% at $\sim$ 100 TeV and $\sim$ 2 PeV, respectively. When comparing the reconstruction properties obtained in the 6-year analysis with those obtained here, they differed slightly. A study of the stability of `taupede` revealed the identification efficiency

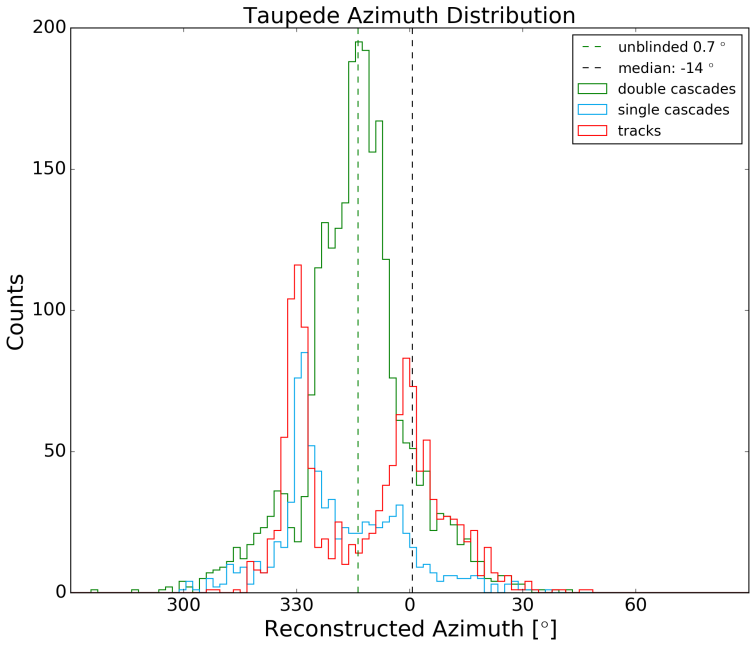

Figure A.2: Distribution of the reconstructed azimuth for the performed seed scan. The seeds resulting in a reclassification of the event show a larger fraction of reconstructed azimuths of the free reconstruction.

of "Double Double" and the stability of the reconstructed observables. A grid of angles was chosen spanning 45° in zenith and 100° in azimuth. In 1° steps, each point was used as a seed for `taupede` to assess the seed dependence. The results are shown in Figure A.1. Green points mark the direction for which the final classification of a double cascade was kept, blue points mark a reclassification to a single cascade, red points a reclassification to a track. 55% of the tried seed directions result in keeping the double cascade classification, 17% of directions result in a reclassification to a single cascade and 28% result in a reclassification as a track.

Figure A.2 shows the reconstructed azimuth of the reconstructions with different seeded directions. For seed directions where the double cascade classification is kept, the distribution is approximately gaussian. The seed directions resulting in a reclassification cluster are peaked at the tails of the double-cascade distribution.

A similar observation can be made when looking at the distribution of reconstructed zenith for the seed directions shown in Figure A.3. While the overall distribution looks approximately gaussian, seeded directions resulting in a reclassification are much more likely to have reconstructed zenith angles below 60° than seeded directions that keep the double cascade classification.

For the vast majority of seeded directions, a high negative energy asymmetry is recovered

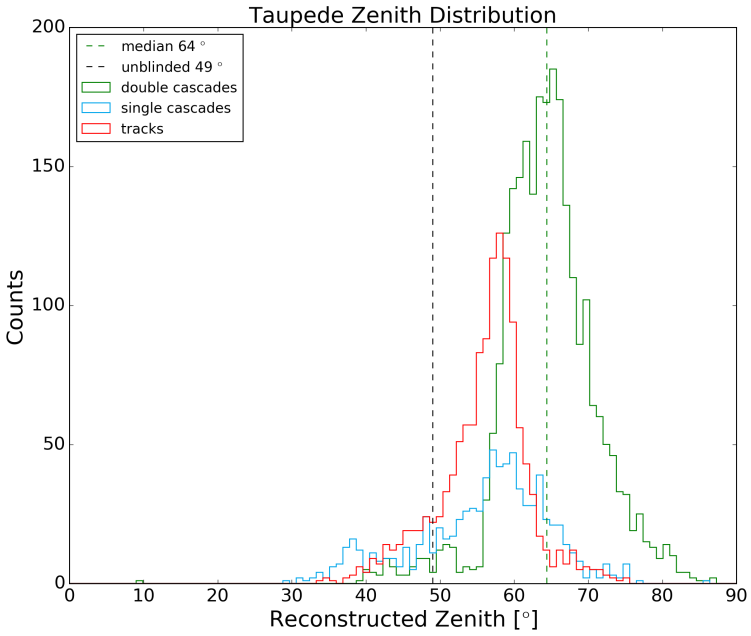

Figure A.3: Distribution of the reconstructed zenith for the performed seed scan. The seeds resulting in a reclassification of the event show a larger fraction of reconstructed zeniths far from the reconstructed zenith of the free reconstruction.

as can be seen in Figure A.4. However, seeded directions resulting in a reclassification show a peak at high positive energy asymmetries.

The double-cascade length distribution is shown in Figure A.5. For seeded directions with kept double-cascade classification the length distribution is approximately gaussian around the free-reconstruction value of 17 m. A small number of seeded directions result in a reconstructed value below 10 m and therefore a reclassification to a single cascade.

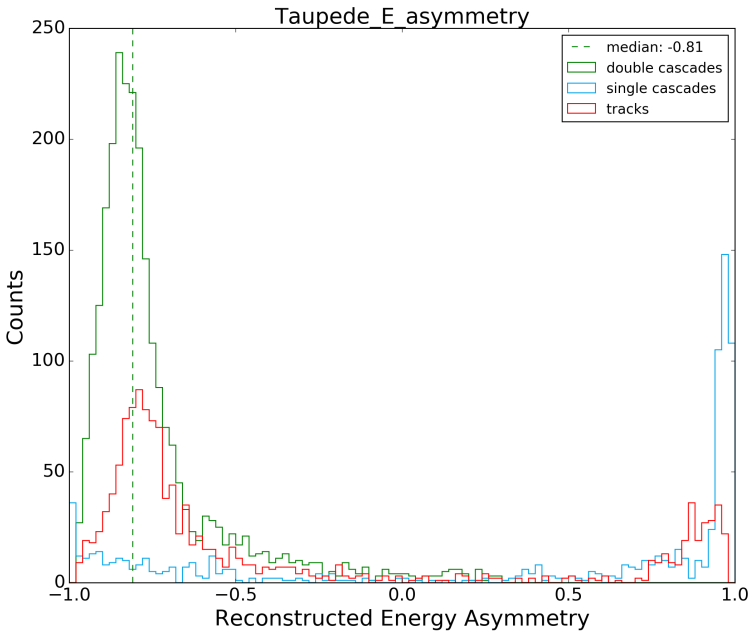

Figure A.4: Distribution of the reconstructed energy asymmetry for the performed seed scan. The seeds resulting in a reclassification show a peak at high positive energy asymmetries, indicative of a bad double-cascade fit.

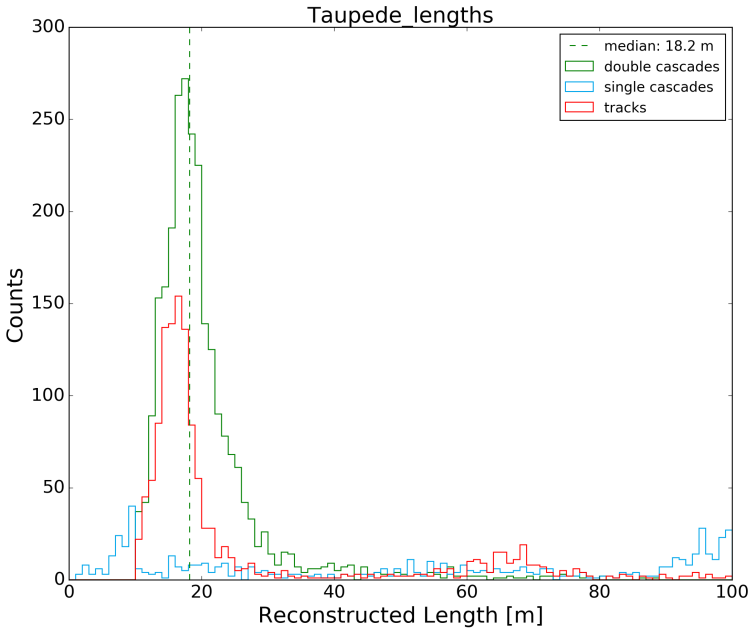

Figure A.5: Distribution of the reconstructed double-cascade length for the performed seed scan. Few seeds resulting in a reclassification as a single cascade have reconstructed length close to the official reconstruction.

Appendix B

Distribution of Reconstructed Properties of the Resimulation Sets

The resimulation procedure for each of the found double cascades is described in Section 7.1. The simulated ranges needed to be determined by trial and error, with the final ranges shown in Table 7.1. While it is possible to estimate reasonable ranges for the primary directions and true interaction vertices based on the reconstruction resolution of IceCube events, it is much more difficult to do so for the primary energy. The reason is that the primary energy is only equal to the deposited, electromagnetic-equivalent energy in the case of ν_e-CC interactions. In all other interactions, some amount of the neutrino energy is invisible, either due to production of neutral or slow particles not emitting Cherenkov light, or due to produced particles leaving the detector. The distributions of all reconstructed properties used in the analysis, and their two-dimensional correlations are shown in Figures B.1, B.2 and B.3 for ν_e-, ν_μ- and ν_τ-resimulation of "Big Bird"; and Figures B.4, B.5 and B.6 for ν_e-, ν_μ- and ν_τ-resimulation of "Double Double". The observed values of the data events are marked with black lines and black stars on the one- and two-dimensional distributions, respectively. The resimulation ranges given in Table 7.1 are clearly visible for all reconstructed properties, except for the reconstructed energy where the ranges imposed on the visible energy losses in the detector can be seen. No significant correlations between any pair of reconstructed properties can be seen. It is notable that for "Big Bird"-like resimulated, very-high-energy events, true single cascades often get reconstructed to double-cascade lengths around 10 m Both events are downgoing, "Big Bird" with a reconstructed zenith angle of 32° and "Double Double" of 48°. As the angular uncertainties are evaluated in degrees, the chosen resimulation

ranges seemed sufficient. Thus, the simulation did not include very down-going events, that could contribute up to $\sim 3\%$ to the signal and $\sim 6\%$ to the overall single cascade background. This was deemed not problematic.

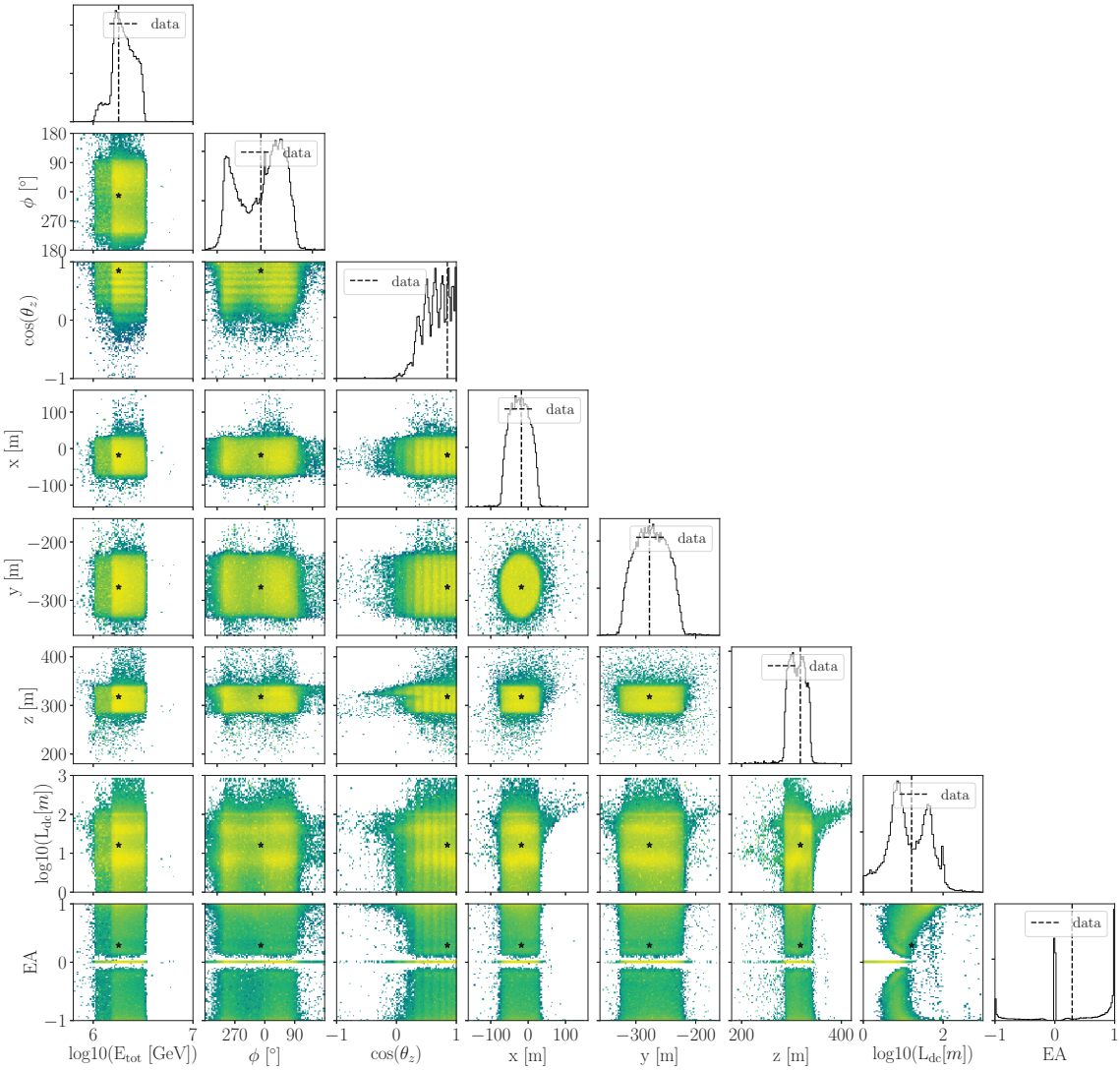

Figure B.1: Distribution of reconstructed properties for "Big Bird"-like resimulated ν_e events passing the HESE selection. One-dimensional histograms for each observable and two-dimensional histograms for each pair of reconstructed properties are shown. "Big Bird's" reconstructed properties are indicated as black lines and black stars on the one- and two-dimensional histograms, respectively.

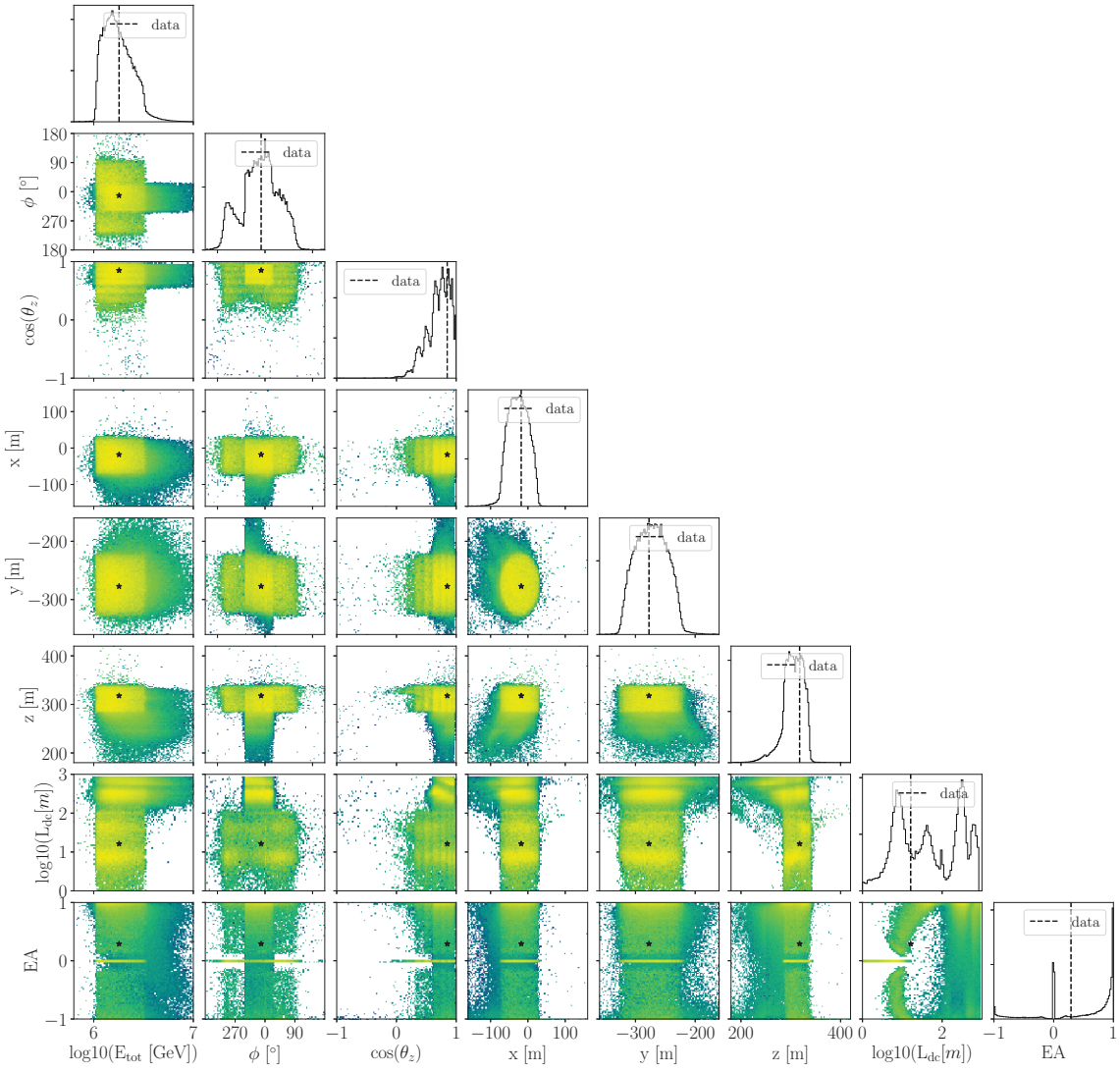

Figure B.2: Distribution of reconstructed properties for "Big Bird"-like resimulated ν_μ events passing the HESE selection. See caption of Figure B.1 for details.

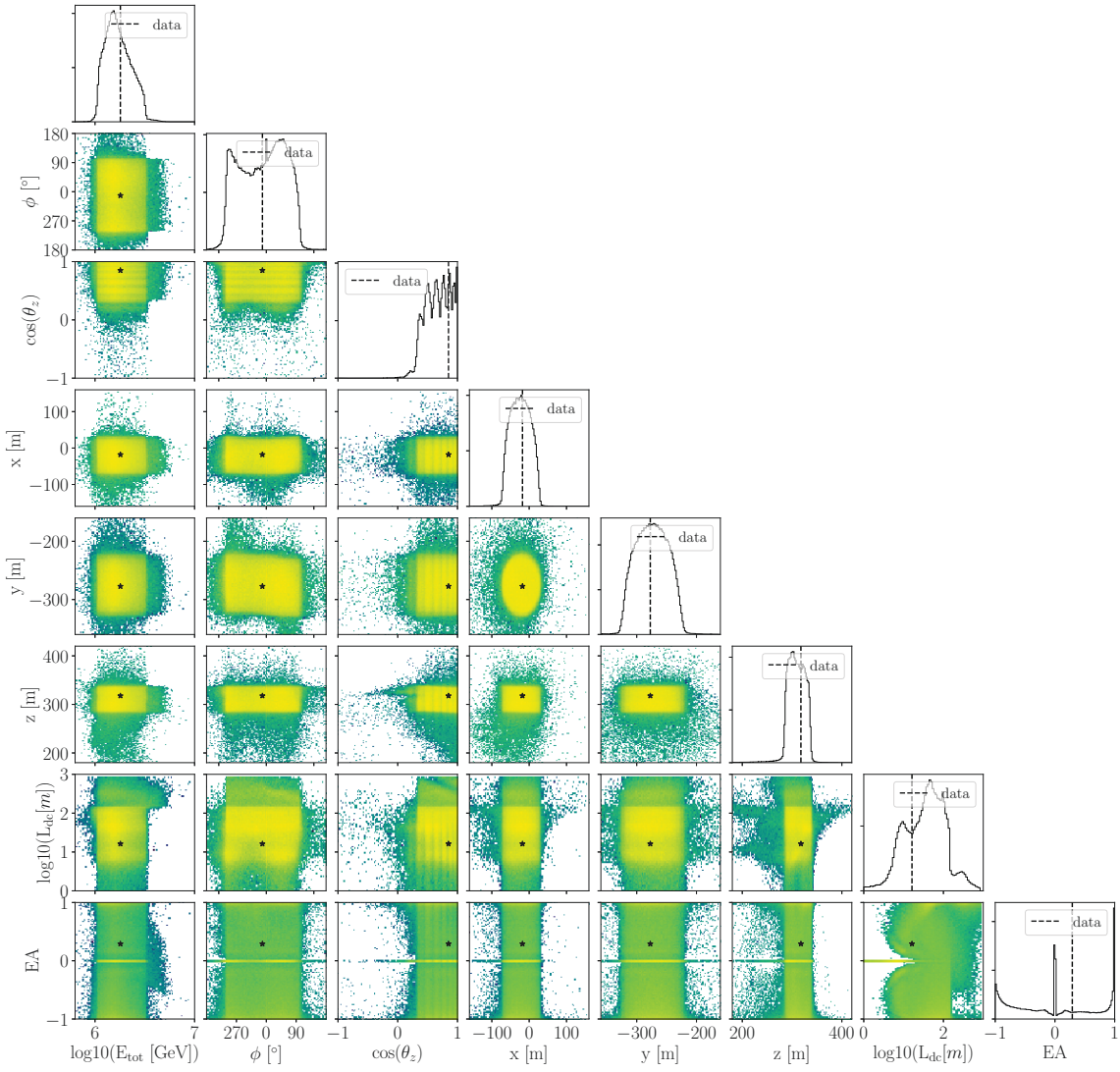

Figure B.3: Distribution of reconstructed properties for "Big Bird"-like resimulated ν_τ events passing the HESE selection. See caption of Figure B.1 for details.

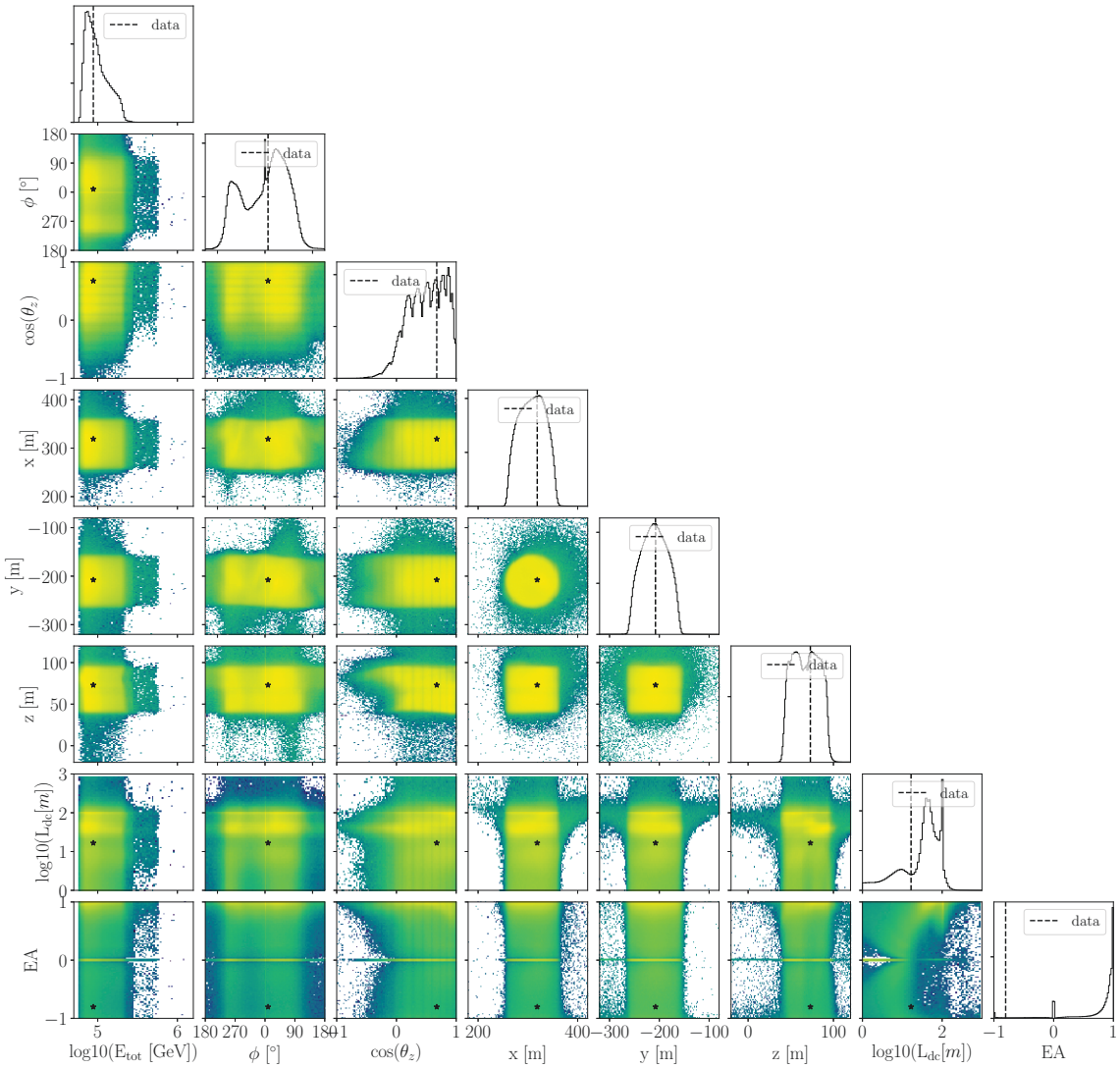

Figure B.4: Distribution of reconstructed properties for "Double Double"-like resimulated ν_e events passing the HESE selection. See caption of Figure B.1 for details.

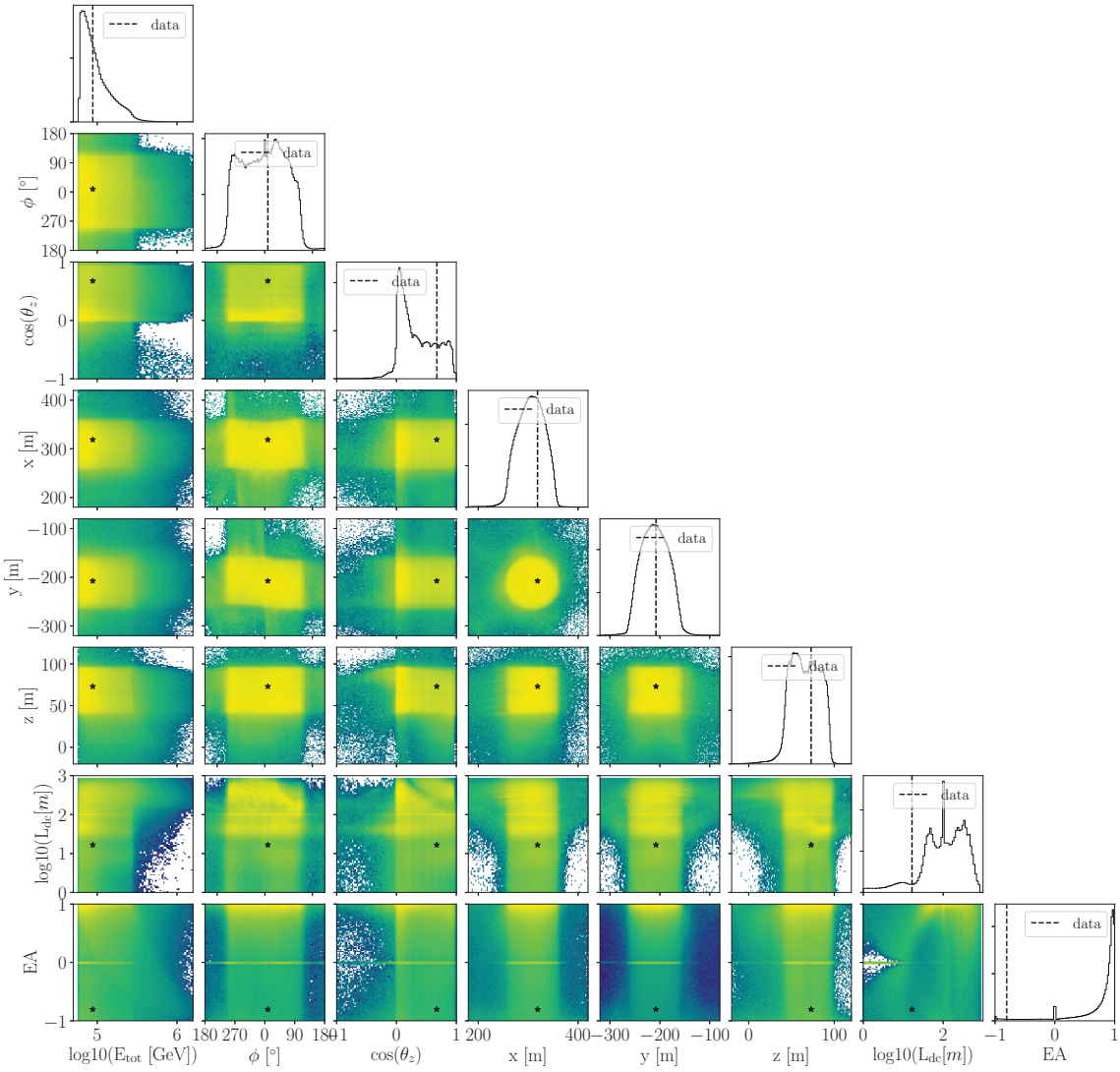

Figure B.5: Distribution of reconstructed properties for "Double Double"-like resimulated ν_μ events passing the HESE selection. See caption of Figure B.1 for details.

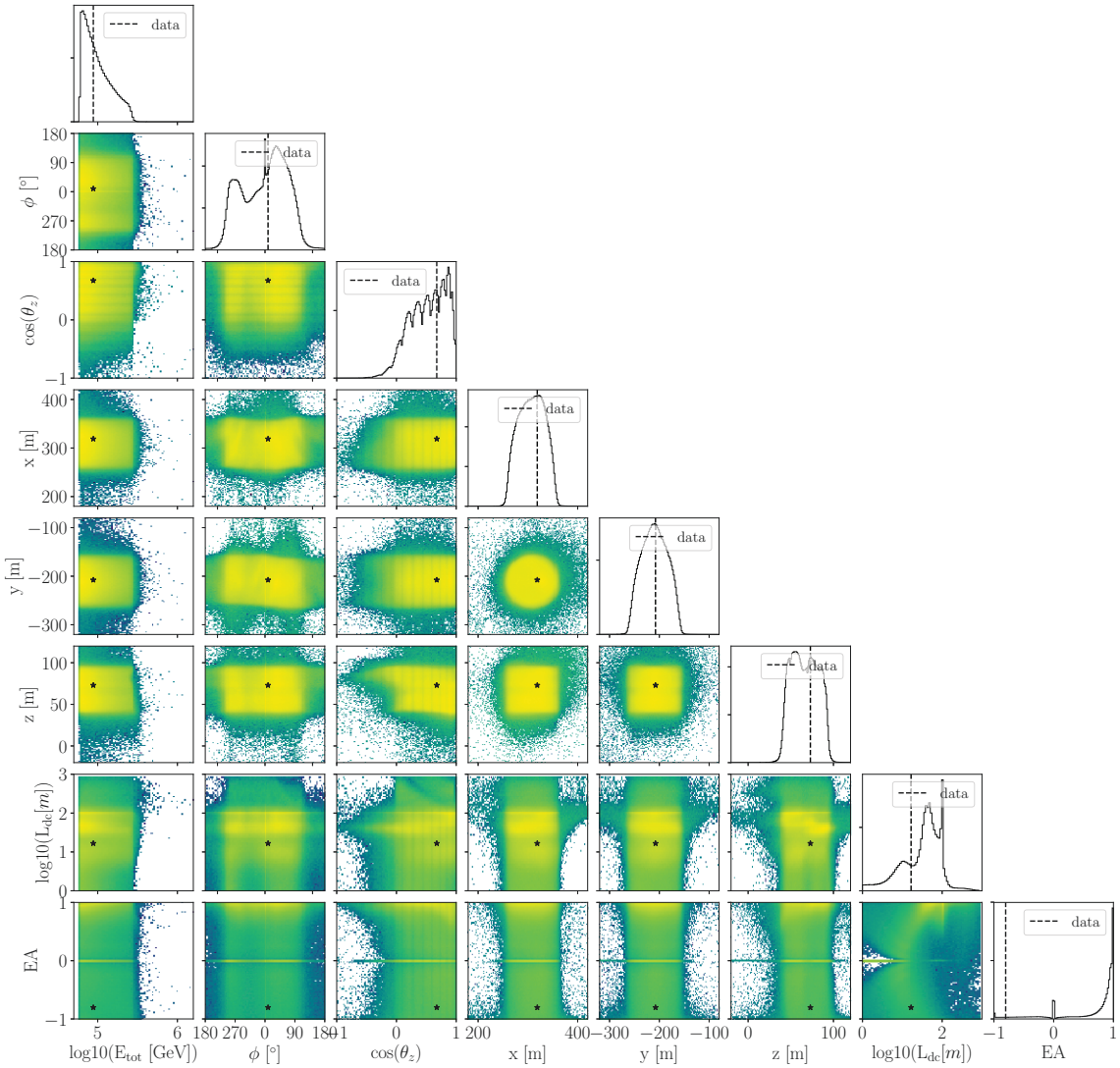

Figure B.6: Distribution of reconstructed properties for "Double Double"-like resimulated ν_τ events passing the HESE selection. See caption of Figure B.1 for details.

Appendix C

On the Validity of Wilks' Theorem

Throughout the presentation of results, confidence regions were drawn assuming Wilks' theorem [151] holds. As described in Section 5.3, Wiks' theorem states: The log-likelihood ratio $-2\Delta \log \mathcal{L} = -2 \log(\mathcal{L}_{\text{test}}/\mathcal{L}_{\text{max}})$ is χ^2_k-distributed, where k is the difference in the number of degrees of freedom of the tested scenario with likelihood $\mathcal{L}_{\text{test}}$ to the best fit with maximum likelihood $\mathcal{L}_{\text{max}}$. However, this holds under the following assumptions:

1. the constrained fit must be a special case of the free fit, i.e. the hypothesis is nested
2. the parameters of interest and the nuisance parameters are not bounded
3. the number of observed events n is large.

The first requirement is fulfilled, namely that the tested hypothesis is a special case of the free fit hypothesis. For the flavor composition measurement presented in Sections 6.1 and 7.3, the tested scenarios constrained the flavor composition parameters, and thus had 2 degrees of freedom than the free single-power-law flavor fit. For the profile likelihood scan of the astrophysical tau neutrino normalization shown in Figure 7.6, the tested scenarios had the tau contribution to the overall astrophysical neutrino flux fixed, and thus had 1 degree of freedom less than the free fit.
The second requirement is clearly not fulfilled: the tau neutrino normalization is bounded by zero. For the flavor composition measurement, the fractions of all neutrino flavors have to sum up to $\sum_{\alpha=e,\mu,\tau} f_\alpha = 1$.
The third requirement is clearly violated as well: We observe two double cascades, and 60 events in total. Neither the double cascade sub-sample, nor the total event sample can be called "large".

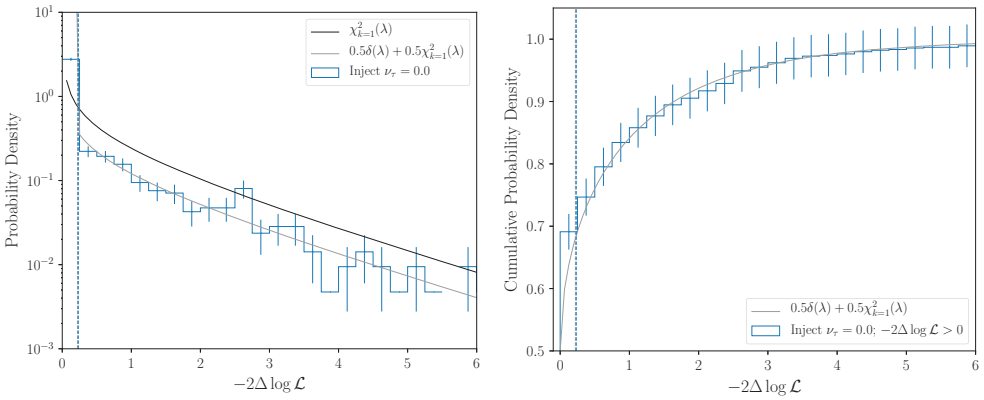

Figure C.1: Test statistic distribution (*left*) and cumulative distribution (*right*) of the one-dimensional pseudo Monte Carlo trials for an injected vanishing astrophysical tau neutrino flux. The expected half-χ^2-distribution (gray line) is followed. The χ^2_1 distribution is shown for reference (black line), but as expected does not describe the generated distribution of pseudo MC experiments well. The vertical lines that mark the events contributing to the 68% confidence interval, and the expectation from the half-χ^2 distribution are almost perfectly aligned.

The validity of Wilks' theorem can be assessed by performing pseudo Monte Carlo experiments. A pseudo Monte Carlo experiment is a realization of the experiment based on Monte Carlo simulated data and the tested model. The tested model is injected as the truth, according to which events are chosen from the simulation as the realization. Thus, the pseudo Monte Carlo experiment can only be performed using the untargeted MC dataset containing all possible realizations, as the realization is chosen randomly according to the injected model. The targeted simulation, however, was generated for the specific realization of the observed double cascades. Therefore, the validity of Wilks' theorem can only be tested for the general case.

For the measurement of the astrophysical tau neutrino flux, 2000 pseudo MC trials are performed by injecting the parameters $\Phi_{\nu_\tau} = 0, \hat{\hat{\theta}}, \hat{\hat{\xi}}$ as the truth, where $\hat{\hat{\theta}}, \hat{\hat{\xi}}$ are the conditional best-fit model parameters. In the case of bounded parameters and a measurement close to the boundary, the approximately followed distribution of the $-2\Delta \log \mathcal{L}$ is a half-chi-squared distribution, $0.5\delta(\lambda) + 0.5\chi^2_{k=1}(\lambda)$ [152]. The distribution of $-2\Delta \log \mathcal{L}$ is obtained by fitting each trial twice: once imposing the injected condition ($\Phi_{\nu_\tau} = 0$), and once leaving all parameters free. The distribution is shown in Figure C.1. The expected half-χ^2-distribution is followed closely. The conclusion can be drawn that for the one-dimensional case of rejecting a vanishing astrophysical tau neutrino flux, Wilks theorem holds and the $-2\Delta \log \mathcal{L}$ can be converted to a confidence level. Even though

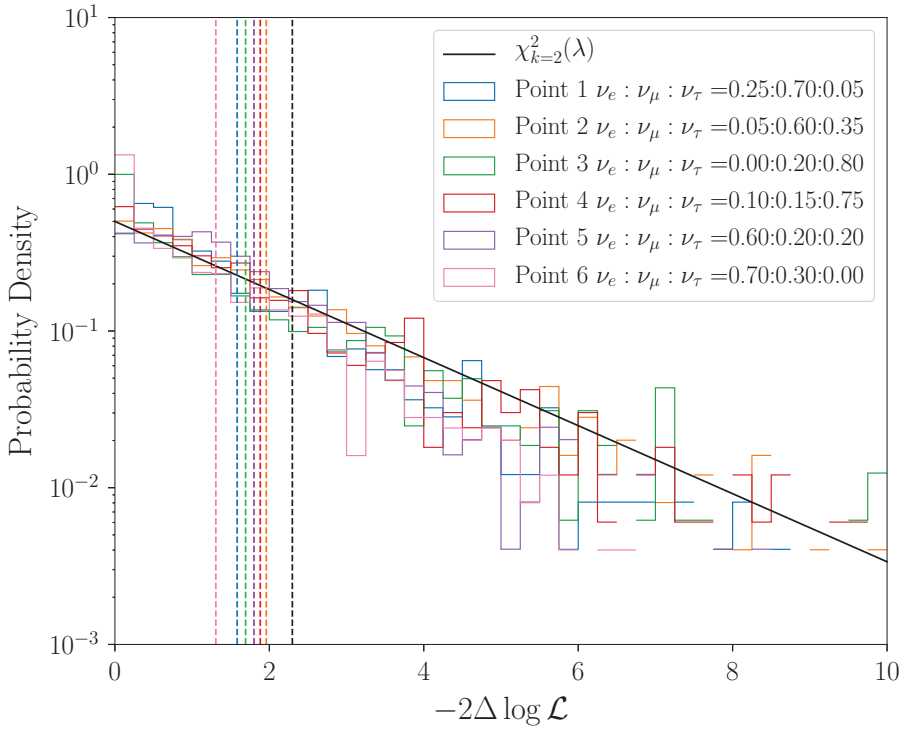

Figure C.2: Test statistic distribution of the two-dimensional pseudo Monte Carlo trials for different evaluated relative flavor fractions.

the same test cannot be performed for the targeted MC case, there is no reason a violation would occur in special parts of the detector and for special reconstructed energy ranges, considering the very good match over the entire detector and the entire energy range.

To test the coverage of the flavor composition measurement, 6 points along the 1σ contour of Figure 6.1 are selected, and pseudo MC trials performed. The injected model is the tested flavor composition and the conditional best-fit values for the remaining parameters. Again, each trial is fitted twice, once with the flavor composition fixed to the injected value, and once leaving all parameters free. The tested flavor compositions and their distributions of $-2\Delta \log \mathcal{L}$ are shown in Figure C.2. As there is a difference of two degrees of freedom between the constrained and the free fit, the distributions are expected to follow a $\chi^2_{k=2}$ distribution, at least at the points where the composition is non-zero in all flavors. The $-2\Delta \log \mathcal{L}$ distributions can be reasonably well described by a $\chi^2_{k=2}$ distribution. A perfect description should not be expected: the two fitting parameters are not completely independent, thus the number of degrees of freedom may be slightly lower than 2.

To conclude, Wilks' theorem appears to hold over the entire parameter space. It should therefore also hold in the special parameter space realized in nature, and can be applied to both the one- and two-dimensional parameter scan to obtain the confidence intervals presented in Section 7.3.

List of Abbreviations

AGN	—	Active Galactic Nucleus
AMANDA	—	Antarctic Muon And Neutrino Detector Array
ANTARES	—	Astronomy With A Neutrino Telescope And Abyss Environmental Research
ATWD	—	Analog Transient Waveform Digitizer
BSM	—	Beyond the Standard Model
CMB	—	Cosmic Microwave Background
CR	—	Cosmic Ray
DAQ	—	Data Acquisition System
DIS	—	Deep Inelastic Scattering
DOM	—	Digital Optical Module
DUMAND	—	Deep Underwater Muon And Neutrino Detector
EeV	—	Exa electron Volt, 10^{18} eV
fADC	—	fast Analog to Digital Converter
GR	—	Glashow Resonance
GRB	—	Gamma-Ray-Burst
GZK	—	Greisen-Zatsepin-Kuzmin
HESE	—	High-Energy Starting Events
HLC	—	Hard Local Coincidence
ICL	—	IceCube Laboratory
IO	—	Inverted Ordering
LED	—	Light-Emitting Diode
MESE	—	Medium-Energy Starting Events
MIM	—	Minimum-Ionizing Muon
MC	—	Monte Carlo
NO	—	Normal Ordering

PDF	—	Probability Distribution Function
PE	—	Photoelectron
PEPE	—	PeV Energy Partially-contained Events
PeV	—	Peta electron Volt, 10^{15} eV
PMNS	—	Pontecorvo-Maki-Nakagawa-Sakata
PMT	—	Photomultiplier Tube
RODEO	—	Regularization Of Derivative Expectation Operator
ROI	—	Region of Interest
SM	—	Standard Model
SNR	—	Supernova Remnant
TDE	—	Tidal Disruption Event
TeV	—	Tera electron Volt, 10^{12} eV
TGM	—	Through-Going Muons
UHE	—	Ultrahigh-Energy

List of Figures

List of Tables

Acknowledgements

First of all I want to thank Marek Kowalski for giving me the opportunity to come to DESY and conduct my research in the IceCube group. I am very grateful for having Markus Ackermann as my adviser, thank you for all the assistance, time, discussions and ideas on yet another way to solve a problem I was stuck on, and for giving me the freedom to work where, when and with whom I needed to.
Thank you to Joanna Kiryluk for introducing me to IceCube and high-energy neutrinos.
A special thank you goes to Albrecht Karle and Queen Kim for making my stays in Madison possible. To the HESE-7 taskforce – Nancy Wandkowsky, Austin Schneider, Carlos Arguelles, Tianlu Yuan, Hrvoje Dujmovic, and Shivesh Mandalia –, thank you for pushing through this behemoth of analyses together, and creating something that's more than the sum of its pieces. For welcoming me in Madison, for pushing through late-night and early-morning calls, for never letting me sleep.
I want to thank my longtime office mates Jakob van Santen an Marcel Usner – I have learned so much from you about reconstruction and analysis techniques, and taus. Thank you to Marcel for developing the flavor composition analysis that I took over.
I am very grateful to Federica Bradascio, Summer Blot, Jakob van Santen, Wing Yan Ma, Simone Garrapa, Christoph Welling, Daniel Fernandez and Robert Stein for proofreading parts of this thesis, and Markus Ackermann for proofreading all of it.
Thank you to Walter Winter and Claudio Kopper for reviewing this thesis.
Thank you fellow students and postdocs at DESY for coffee time, the Friday afternoon train beers, dinners, parties; for becoming friends. Thank you everyone at DESY for creating such a nice and welcoming atmosphere .
A very special thank you goes to my parents, who have always supported me in every decision I made, dziękuję, kochani. Tyler, thank you for putting up with me.

Bibliography

[1] V. F. Hess, *Phys. Z.* **13** (1912) 1084–1091.

[2] W. Pauli, "Offener brief an die gruppe der radioaktiven bei der gauverein-tagung zu tübingen." https://cds.cern.ch/record/83282/files/meitner_0393.pdf, 1930.

[3] E. Fermi, *Z. Phys.* **88** (1934) 161–177.

[4] **Planck** Collaboration, N. Aghanim et al., arXiv:1807.06209.

[5] **KATRIN** Collaboration, M. Aker et al., arXiv:1909.06048.

[6] **Project 8** Collaboration, A. Ashtari Esfahani et al., *J. Phys.* **G44** (2017) 054004.

[7] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, *Science* **124** (1956) 103–104.

[8] G. Danby, J. M. Gaillard, K. A. Goulianos, L. M. Lederman, N. B. Mistry, M. Schwartz, and J. Steinberger, *Phys. Rev. Lett.* **9** (1962) 36–44.

[9] **DONUT** Collaboration, K. Kodama et al., *Phys. Lett.* **B504** (2001) 218–224.

[10] **OPERA** Collaboration, N. Agafonova et al., *Phys. Rev. Lett.* **120** (2018) 211801. [Erratum: Phys. Rev. Lett.121,no.13,139901(2018)].

[11] J. N. Bahcall, *Phys. Rev. Lett.* **12** (1964) 300–302. [,9(1964)].

[12] R. Davis, Jr., D. S. Harmer, and K. C. Hoffman, *Phys. Rev. Lett.* **20** (1968) 1205–1209.

[13] B. Pontecorvo, *Sov. Phys. JETP* **6** (1957) 429. [Zh. Eksp. Teor. Fiz.33,549(1957)].

[14] Z. Maki, M. Nakagawa, and S. Sakata, *Prog.Theor.Phys.* **28** (1962) 870–880.

[15] C. Spiering, arXiv:1903.11481 in *International Conference on History of the Neutrino: 1930-2018 Paris, France, September 5-7, 2018*, (2019).

[16] M. A. Markov and I. M. Zheleznykh, *Nucl. Phys.* **27** (1961) 385–394.

[17] **DUMAND** Collaboration, J. Babson et al., *Phys. Rev.* **D42** (1990) 3613–3620.

[18] **BAIKAL** Collaboration, I. A. Belolaptikov et al., *Astropart. Phys.* **7** (1997) 263–282.

[19] **Baikal** Collaboration, V. A. Balkanov et al., *Astropart. Phys.* **12** (1999) 75–86.

[20] F. Halzen and J. G. Learned.

[21] E. Andres et al., *Astropart. Phys.* **13** (2000) 1–20.

[22] **IceCube** Collaboration, M. G. Aartsen et al., *JINST* **12** (2017) P03012.

[23] **IceCube** Collaboration, M. G. Aartsen et al., *Science* **342** (2013).

[24] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev. Lett.* **115** (2015) 081102.

[25] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev.* **D91** (2015) 022001.

[26] **IceCube** Collaboration, H. M. Niederhausen, M. Lesiak-Bzdak, and A. Stoessl, PoS(ICRC2015)1109 (2016).

[27] **Kamiokande-II** Collaboration, K. Hirata et al., *Phys. Rev. Lett.* **58** (1987) 1490–1493. [,727(1987)].

[28] R. M. Bionta et al., *Phys. Rev. Lett.* **58** (1987) 1494.

[29] E. N. Alekseev, L. N. Alekseeva, I. V. Krivosheina, and V. I. Volchenko, *Phys. Lett.* **B205** (1988) 209–214.

[30] I. Esteban, M. C. Gonzalez-Garcia, A. Hernandez-Cabezudo, M. Maltoni, and T. Schwetz, *JHEP* **01** (2019) 106.

[31] R. W. Rasmussen, L. Lechner, M. Ackermann, M. Kowalski, and W. Winter, *Phys. Rev.* **D96** (2017) 083018.

[32] **IceCube** Collaboration, M. G. Aartsen et al., *Astrophys. J.* **809** (2015) 98.

[33] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev. Lett.* **114** (2015) 171102.

[34] **IceCube** Collaboration, R. Abbasi et al., *Phys. Rev.* **D86** (2012) 022005.

[35] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev.* **D93** (2016) 022001.

[36] Patrick Hallen, *On the Measurement of High-Energy Tau Neutrinos with IceCube*, Master's thesis, RWTH Aachen, 2013.

[37] **IceCube** Collaboration, M. Usner et al. in *Proceedings, 35th International Cosmic Ray Conference (ICRC 2017): Bexco, Busan, Korea, 10-20 July, 2017*, vol. ICRC2017, p. 974, (2018).

[38] **Particle Data Group** Collaboration, M. Tanabashi et al., *Phys. Rev.* **D98** (2018) 030001.

[39] T. K. Gaisser, *Cosmic rays and particle physics*. Cambridge, UK: Univ. Pr. (1990) 279 p, 1990.

[40] K. Greisen, *Phys. Rev. Lett.* **16** (1966) 748–750.

[41] G. T. Zatsepin and V. A. Kuzmin, *JETP Lett.* **4** (1966) 78–80. [Pisma Zh. Eksp. Teor. Fiz.4,114(1966)].

[42] **Pierre Auger** Collaboration, PoS(ICRC2019)440 (2020).

[43] **Pierre Auger** Collaboration, PoS(ICRC2019)482 (2020).

[44] **Pierre Auger** Collaboration, A. Aab et al., *Nucl. Instrum. Meth.* **A798** (2015) 172–213.

[45] **Telescope Array** Collaboration, T. Abu-Zayyad et al., *Nucl. Instrum. Meth.* **A689** (2013) 87–97.

[46] A. Fedynitch, R. Engel, T. K. Gaisser, F. Riehn, and T. Stanev, *EPJ Web Conf.* **99** (2015) 08001.

[47] **Pierre Auger** Collaboration, PoS(ICRC2019)404 (2020).

[48] **Pierre Auger** Collaboration, PoS(ICRC2019)411 (2020).

[49] **IceCube** Collaboration, M. G. Aartsen et al., *Astrophys. J.* **849** (2017) 67.

[50] **IceCube** Collaboration, M. G. Aartsen et al., *Astrophys. J.* **886** (2019) 12.

[51] **IceCube** Collaboration, M. G. Aartsen et al., *Astrophys. J.* **843** (2017) 112.

[52] **IceCube** Collaboration, M. G. Aartsen et al., *Astrophys. J.* **835** (2017) 45.

[53] A. J. Stasik. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2018.

[54] **IceCube** Collaboration, R. Stein et al., arXiv:1908.08547 in *Proceedings, 36th International Cosmic Ray Conference (ICRC 2019): Madison, WI, USA, July 24-August 1, 2019*, (2019).

[55] **IceCube** Collaboration, T. Carver et al., arXiv:1908.05993 in *Proceedings, 36th International Cosmic Ray Conference (ICRC 2019): Madison, WI, USA, July 24-August 1, 2019*, (2019).

[56] **IceCube** Collaboration, M. G. Aartsen et al., *Astropart. Phys.* **92** (2017) 30–41.

[57] **IceCube** Collaboration, C. Kopper and E. Blaufuss, *GCN Circular* (2017).

[58] **Fermi-LAT** Collaboration, D. K. Yasuyuki T. Tanaka, Sara Buson, *Astronomer's Telegram* (2017).

[59] **MAGIC** Collaboration, R. Mirzoyan, *Astronomer's Telegram* (2017).

[60] **IceCube, Fermi-LAT, MAGIC, AGILE, ASAS-SN, HAWC, H.E.S.S., INTEGRAL, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, Swift NuSTAR, VERITAS, VLA/17B-403** Collaboration, M. G. Aartsen et al., *Science* **361** (2018) eaat1378.

[61] **IceCube** Collaboration, M. G. Aartsen et al., *Science* **361** (2018) 147–151.

[62] W. Winter et al., PoS(ICRC2019)1032 (2020).

[63] L. Landau and E. Lifshitz, *Fluid Mechanics*. Pergamon, second edition ed., 1987.

[64] R. Alves Batista et al., *Front. Astron. Space Sci.* **6** (2019) 23.

[65] A. M. Hillas, *Ann. Rev. Astron. Astrophys.* **22** (1984) 425–444.

[66] T. Kashti and E. Waxman, *Phys. Rev. Lett.* **95** (2005) 181101.

[67] S. Hummer, M. Maltoni, W. Winter, and C. Yaguna, *Astropart. Phys.* **34** (2010) 205–224.

[68] L. A. Anchordoqui, H. Goldberg, F. Halzen, and T. J. Weiler, *Phys. Lett.* **B593** (2004) 42.

[69] S. L. Glashow, *Phys. Rev.* **118** (1960) 316–317.

[70] M. Kachelriess and R. Tomas, *Phys. Rev.* **D74** (2006) 063009.

[71] M. Ahlers, L. A. Anchordoqui, M. C. Gonzalez-Garcia, F. Halzen, and S. Sarkar, *Astropart. Phys.* **34** (2010) 106–115.

[72] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev.* **D98** (2018) 062003.

[73] I. Safa, A. Pizzuto, C. Arguelles, F. Halzen, R. Hussain, A. Kheirandish and J. Vandenbroucke in *Proceedings, 36th International Cosmic Ray Conference (ICRC 2019): Madison, WI, USA, July 24-August 1, 2019*, (2019).

[74] R. Davis, *Phys. Rev. Lett.* **12** (1964) 303–305. [107(1964)].

[75] **Super-Kamiokande** Collaboration, Y. Fukuda et al., *Phys. Rev. Lett.* **81** (1998) 1562–1567.

[76] **SNO** Collaboration, Q. R. Ahmad et al., *Phys. Rev. Lett.* **89** (2002) 011301.

[77] **SNO** Collaboration, Q. R. Ahmad et al., *Phys. Rev. Lett.* **89** (2002) 011302.

[78] S. Mikheev and A. Y. Smirnov, *Nuovo Cim.* **C9** (1986) 17–26.

[79] L. Wolfenstein, *Phys.Rev.* **D17** (1978) 2369–2374.

[80] M. Gell-Mann, P. Ramond, and R. Slansky, *Conf. Proc.* **C790927** (1979) 315–321.

[81] T. Yanagida, *Progress of Theoretical Physics* **64** (Sep, 1980) 1103–1105.

[82] R. N. Mohapatra and G. Senjanovic, *Phys. Rev. Lett.* **44** (Apr, 1980) 912–915.

[83] "NuFit4.1." www.nu-fit.org, 2019.

[84] **IceCube** Collaboration, M. G. Aartsen et al., *in preparation*.

[85] L. D. Landau and I. Pomeranchuk, *Dokl. Akad. Nauk Ser. Fiz.* **92** (1953) 535–536.

[86] A. B. Migdal, *Phys. Rev.* **103** (1956) 1811–1820.

[87] **Particle Data Group** Collaboration, C. Patrignani et al., *Chin. Phys.* **C40** (2016) 100001.

[88] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev. Lett.* **113** (2014) 101101.

[89] **IceCube** Collaboration, PoS(ICRC2019)1017 (2020).

[90] **IceCube** Collaboration, H. M. "Niederhausen and Y. Xu, PoS(ICRC2017)968 (2018).

[91] **IceCube** Collaboration, M. G. Aartsen et al., *in preparation*.

[92] **IceCube** Collaboration, L. Lu, PoS(ICRC2017)1002 (2018).

[93] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev. Lett.* **120** (2018) 071801.

[94] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev.* **D99** (2019) 032007.

[95] "IceCube gallery." https://gallery.icecube.wisc.edu/internal/main.php.

[96] U. F. Katz, *Nucl. Instrum. Meth.* **A567** (2006) 457–461.

[97] **ANTARES** Collaboration, J. A. Aguilar et al., *Astropart. Phys.* **26** (2006) 314–324.

[98] **Baikal-GVD** Collaboration, A. D. Avrorin et al., arXiv:1908.05427 in *36th International Cosmic Ray Conference (ICRC 2019) Madison, Wisconsin, USA, July 24-August 1, 2019*, (2019).

[99] T. Benson, J. Cherwinka, M. Duvernois, A. Elcheikh, F. Feyzi, L. Greenler, J. Haugen, A. Karle, M. Mulligan, and R. Paulos, *Annals of Glaciology* **55** (2014) 105–114.

[100] **IceCube** Collaboration, R. Abbasi and et al., *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **618** (Jun, 2010) 139–152.

[101] **IceCube** Collaboration, R. Abbasi and et al., *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **601** (Apr, 2009) 294–316.

[102] **IceCube** Collaboration, R. Cross, A. Fritz, and S. Griswold, arXiv:1908.07249 in *Proceedings, 35th International Cosmic Ray Conference (ICRC 2017): Bexco, Busan, Korea, 10-20 July, 2017*, (2019).

[103] **IceCube** Collaboration, M. G. Aartsen et al., *in preparation*.

[104] **IceCube** Collaboration, M. G. Aartsen et al., *in preparation*.

[105] **IceCube** Collaboration, M. G. Aartsen et al., *JINST* **9** (2014) P03009.

[106] M. Usner. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2018.

[107] M. Ackermann et al., *J. Geophys. Res.* **111** (2006) D13203.

[108] R. C. Bay, "IceCube Dust Map." http://icecube.berkeley.edu/~bay/dustmap/.

[109] D. Chirkin, "Spice3.2 ice model." https://events.icecube.wisc.edu/event/75/contributions/4869/attachments/3947/4325/icetalk.pdf.

[110] **IceCube** Collaboration, M. G. Aartsen et al., *Nucl. Instrum. Meth.* **A711** (2013) 73–89.

[111] **IceCube** Collaboration, D. Chirkin in *"Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013"*, p. 0580, (2014).

[112] **IceCube** Collaboration, D. Chirkin and M. Rongen, arXiv:1908.07608 in *Proceedings, 36th International Cosmic Ray Conference (ICRC 2019): Madison, WI, USA, July 24-August 1, 2019*, (2019).

[113] R. C. Bay, R. A. Rohde, P. B. Price, and N. E. Bramall, *Journal of Geophysical Research: Atmospheres* **115** (2010).

[114] **IceCube** Collaboration, M. G. Aartsen et al., *Nucl. Instrum. Meth.* **A711** (2013) 73–89.

[115] P. A. Cerenkov, *Phys. Rev.* **52** (1937) 378–379.

[116] M. P. Kowalski. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I, 2004.

[117] A. Gazizov and M. P. Kowalski, *Comput. Phys. Commun.* **172** (2005) 203–213.

[118] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, *CORSIKA: a Monte Carlo code to simulate extensive air showers.* Feb., 1998.

[119] J. van Santen. PhD thesis, University of Wisconsin-Madison, 2014.

[120] J. H. Koehne, K. Frantzen, M. Schmitz, T. Fuchs, W. Rhode, D. Chirkin, and J. Becker Tjus, *Comput. Phys. Commun.* **184** (2013) 2070–2090.

[121] B. Voigt. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2008.

[122] C. Kopper, "CLSIM." https://github.com/claudiok/clsim, 2012.

[123] N. Whitehorn, J. van Santen, and S. Lafebre, *Comput. Phys. Commun.* **184** (2013) 2214–2220.

[124] J. van Santen. private communication.

[125] **AMANDA** Collaboration, J. Ahrens et al., *Nucl. Instrum. Meth.* **A524** (2004) 169–194.

[126] Dirk Pandel, *Bestimmung von Wasser- und Detektorparametern und Rekonstruktion von Myonen bis 100 TeV mit dem Baikal-Neutrinoteleskop NT-72*, 1996.

[127] **AMANDA** Collaboration, R. Abbasi et al., *Astropart. Phys.* **34** (2011) 420–430.

[128] R. Gandhi, C. Quigg, M. H. Reno, and I. Sarcevic, *Astropart. Phys.* **5** (1996) 81–110.

[129] F. Halzen and D. Saltzberg, *Phys. Rev. Lett.* **81** (1998) 4305–4308.

[130] J. G. Learned and S. Pakvasa, *Astropart. Phys.* **3** (1995) 267–274.

[131] **IceCube** Collaboration, PoS(ICRC2015)1081 (2016).

[132] **IceCube** Collaboration, PoS(ICRC2017)981 (2018).

[133] C. A. Argüelles, A. Schneider and T. Yuan, *Journal of High Energy Physics* **2019** (2019) 30.

[134] "GOLEM." https://github.com/IceCubeOpenSource/GolemFit.

[135] **IceCube** Collaboration, N. Wandkowsky and C. Weaver in *Proceedings, 35th International Cosmic Ray Conference (ICRC 2017): Bexco, Busan, Korea, 10-20 July, 2017*, vol. ICRC2017, p. 976, (2018).

[136] C. A. Argüelles, S. Palomares-Ruiz, A. Schneider, L. Wille, and T. Yuan, *JCAP* **1807** (2018) 047.

[137] T. K. Gaisser, K. Jero, A. Karle, and J. van Santen, *Phys. Rev.* **D90** (2014) 023009.

[138] A. Fedynitch, "MCEq." https://github.com/afedynitch/MCEq, 2017.

[139] T. K. Gaisser, *Astropart. Phys.* **35** (2012) 801–806.

[140] F. Riehn, H. P. Dembinski, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, PoS(ICRC2017)301 (2018). [35,301(2017)].

[141] U. von Zahn, *Journal of Atmospheric and Terrestrial Physics* **49** (1987) 607 – 620. The Middle Atmosphere Program/Winter in Northern Europe Project.

[142] A. E. Hedin, *Journal of Geophysical Research: Space Physics* **96** (1991) 1159–1172.

[143] **IceCube** Collaboration, M. G. Aartsen et al., *in preparation*.

[144] **IceCube** Collaboration, M. G. Aartsen et al., *in preparation*.

[145] **IceCube** Collaboration, M. G. Aartsen et al., *in preparation*.

[146] **IceCube** Collaboration, N. Wandkowsky, C. Weaver, et al. in *Proceedings, 35th International Cosmic Ray Conference (ICRC 2017): Bexco, Busan, Korea, 10-20 July, 2017*, vol. ICRC2017, p. 976, (2018).

[147] M. Honda, T. Kajita, K. Kasahara, S. Midorikawa and T. Sanuki, *Phys. Rev. D* **75** (2007) 043006.

[148] A. Bhattacharya, R. Enberg, M. H. Reno, I. Sarcevic, and A. Stasto, *JHEP* **06** (2015) 110.

[149] T. Montaruli and F. Ronga, arXiv:1109.6238.

[150] T. K. Gaisser, T. Stanev, and S. Tilav, *Front. Phys.(Beijing)* **8** (2013) 748–758.

[151] S. S. Wilks, *Annals Math. Statist.* **9** (1938) 60–62.

[152] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Eur. Phys. J.* **C71** (2011) 1554. [Erratum: Eur. Phys. J.C73,2501(2013)].

[153] "Steamshovel." Documentation at: http://software.icecube.wisc.edu/documentation/projects/steamshovel/index.html.

[154] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev.* **D99** (2019) 032004.

[155] P. F. de Salas, R. A. Lineros, and M. Tortola, *Phys. Rev.* **D94** (2016) 123001.

[156] **IceCube** Collaboration, M. Meier and J. Soedingrekso, arXiv:1909.05127 in *Proceedings, 36th International Cosmic Ray Conference (ICRC 2019): Madison, WI, USA, July 24-August 1, 2019*, (2019).

[157] **IceCube** Collaboration, L. Wille and D. Xu, arXiv:1909.05162 in *Proceedings, 36th International Cosmic Ray Conference (ICRC 2019): Madison, WI, USA, July 24-August 1, 2019*, (2019).

[158] **IceCube** Collaboration, J. Stachurska et al. in *Proceedings, 35th International Cosmic Ray Conference (ICRC 2017): Bexco, Busan, Korea, 10-20 July, 2017*, vol. ICRC2017, p. 973, (2018).

[159] **IceCube** Collaboration, A. Ishihara et al., arXiv:1908.09441 in *Proceedings, 36th International Cosmic Ray Conference (ICRC 2019): Madison, WI, USA, July 24-August 1, 2019*, (2019).

[160] **IceCube-Gen2** Collaboration, M. G. Aartsen et al., *in preparation.*

[161] Fuyudi Zhang, *High-Energy Tau Neutrinos with IceCube-Gen2*, Master's thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, in preparation.

Selbstständigkeitserklärung

Ich erkläre, dass ich die Dissertation selbständig und nur unter Verwendung der von mir gemäß § 7 Abs. 3 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät, veröffentlicht im Amtlichen Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 42/2018 am 11.07.2018, angegebenen Hilfsmittel angefertigt habe.

Berlin, den ...,

(Unterschrift)

Ich erkläre, dass die von mir in der Universitätsbibliothek abgegebene schriftliche und elektronische Version der Dissertationsschrift mit der angenommenen Dissertation übereinstimmt.